

# Modelos de sobrevivência aplicados à análise de acontecimentos múltiplos

Ivo Sousa-Ferreira

Departamento de Matemática, Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira e Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal, *ivo.ferreira@staff.uma.pt*

Ana Maria Abreu

Departamento de Matemática, Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira e Centro de Investigação em Matemática e Aplicações, *abreu@staff.uma.pt*

Cristina Rocha

Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal e Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal, *cmrocha@fc.ul.pt*

**Palavras-chave:** Acontecimentos múltiplos; Análise de sobrevivência; Extensões do modelo de Cox; Modelos paramétricos; *Software R*.

**Resumo:** Na modelação de acontecimentos múltiplos, uma abordagem muito utilizada consiste em desenvolver extensões do modelo semiparamétrico de Cox. Contudo, quando se considera que o conhecimento da distribuição do tempo é importante para o estudo, a abordagem paramétrica revela-se mais adequada. Neste trabalho, são apresentados dois novos modelos paramétricos baseados na distribuição de Weibull (modelos WNE e WE), como alternativas a duas das extensões do modelo de Cox para acontecimentos recorrentes (modelos AG e PWP). De forma a ilustrar as diferenças entre estas duas abordagens, é considerado um exemplo de aplicação com

dados simulados. Os resultados obtidos com os modelos propostos permitem acrescentar mais duas opções ao leque de escolhas de modelos para acontecimentos recorrentes.

## 1 Introdução

Em estudos de longa duração é cada vez mais frequente observar a ocorrência de vários acontecimentos para o mesmo indivíduo. Situações deste tipo surgem, por exemplo, nas ciências da saúde, quando se pretende analisar o tempo até sucessivas recaídas de uma doença; e nas ciências económicas, onde interessa avaliar as razões que levam à observação de insolvências de instituições bancárias num determinado país. Assim, tem-se verificado um crescente empenho em desenvolver metodologia estatística capaz de dar resposta a estes cenários complexos.

Na literatura, existem várias abordagens para modelar o tempo até à ocorrência de acontecimentos múltiplos [5]. Aquela que mais tem sido aplicada, dada a sua versatilidade, consiste em adaptar o modelo semiparamétrico de Cox [2] a esta situação. De facto, nas últimas quatro décadas têm sido propostas várias extensões deste modelo que visam ter em conta diversos aspetos relevantes num estudo desta natureza [3, 16], nomeadamente: acontecimentos do mesmo tipo (designados por acontecimentos recorrentes) ou de tipos diferentes; acontecimentos instantâneos ou duradouros; acontecimentos com riscos de ocorrência distintos; acontecimentos com uma estrutura de ordenação; e acontecimentos com uma estrutura de dependência.

Uma outra abordagem, ainda pouco trabalhada no contexto dos acontecimentos múltiplos, consiste na modelação totalmente paramétrica dos tempos até à ocorrência dos acontecimentos. Note-se que, nos modelos que são extensões do modelo de Cox, a função de risco subjacente não é especificada, o que pode constituir uma limitação. Com efeito, em certas circunstâncias, a estimação desta função é extremamente importante, em particular na área da saúde,

pois permite estudar a evolução de uma doença ao longo do tempo. Deste modo, os modelos paramétricos têm a vantagem de permitir a estimação da função de risco subjacente.

O principal propósito deste trabalho é dar a conhecer dois tipos de abordagem para a modelação de acontecimentos recorrentes: semi-paramétrica *versus* paramétrica. Para tal, começa-se por referir a abordagem semiparamétrica, onde serão consideradas duas extensões do modelo de Cox e, em seguida, a abordagem paramétrica, onde serão apresentados dois novos modelos baseados na distribuição de Weibull. Por último, será apresentado um exemplo de aplicação com dados simulados de forma a ilustrar as diferenças entre estas abordagens.

## 2 Metodologia

Os modelos de riscos proporcionais constituem uma classe de modelos de regressão bem conhecida em Análise de Sobrevida. Este tipo de modelos é caracterizado pela proporcionalidade entre as funções de risco respeitantes a indivíduos com diferentes valores das covariáveis. De facto, considerando  $T$  uma variável aleatória (v.a.) contínua e  $\mathbf{z} = (z_1, \dots, z_p)'$  um vetor de covariáveis associado a um determinado indivíduo, o modelo de riscos proporcionais formulado com base na função de risco é usualmente escrito na forma

$$h(t; \mathbf{z}) = h_0(t) \exp(\beta' \mathbf{z}), \quad t \geq 0, \quad (1)$$

em que  $\beta' = (\beta_1, \dots, \beta_p)$  é o vetor de parâmetros de regressão que representa o efeito (desconhecido) das  $p$  covariáveis, e  $h_0(t)$  denota uma função não negativa, designada por função de risco subjacente. Quando não se especifica uma forma particular para esta função, isto é, quando  $h_0(t)$  é uma função arbitrária, obtém-se o modelo semiparamétrico proposto por Sir David Cox [2]. Por outro lado, quando se admite que o tempo de vida segue uma determinada distribuição, o modelo resultante é totalmente paramétrico.

Antes de proceder à formulação dos modelos que serão considerados neste trabalho, interessa introduzir alguma notação adicional. Seja  $n$  o número de indivíduos em estudo, em que para cada um pode ser observado um máximo de  $K$  acontecimentos, e  $X_{ik} = \min\{T_{ik}, C_{ik}\}$  o tempo em observação do indivíduo  $i$  ( $i = 1, \dots, n$ ) correspondente ao acontecimento  $k$  ( $k = 1, \dots, K$ ), onde  $T_{ik}$  e  $C_{ik}$  representam o seu verdadeiro tempo e o seu tempo de censura, respetivamente. A variável indicatriz que caracteriza o estado do  $i$ -ésimo indivíduo em relação ao acontecimento  $k$  é definida por  $\delta_{ik} = I(T_{ik} \leq C_{ik})$ , a qual toma o valor um quando o acontecimento  $k$  é observado e zero caso contrário. Por fim, denote-se por  $\mathbf{z}_{ik}(t) = (z_{ik1}(t), \dots, z_{ikp}(t))'$  o vetor de  $p$  covariáveis (possivelmente dependentes do tempo) associado ao  $i$ -ésimo indivíduo referente ao acontecimento  $k$ .

## 2.1 Extensões do modelo semiparamétrico de Cox

O modelo de Cox é apropriado para analisar o tempo decorrido desde um instante inicial, bem definido, até à observação de um único acontecimento de interesse. Porém, no contexto dos acontecimentos múltiplos, o facto de se poder registar mais do que um tempo para cada indivíduo inviabiliza a aplicação direta deste modelo. Por essa razão, têm sido propostas várias extensões do modelo de Cox para analisar acontecimentos múltiplos e, em particular, acontecimentos de um único tipo que se repetem ao longo do tempo [16]. Este trabalho focar-se-á nesta segunda situação, sendo que duas das extensões do modelo de Cox que mais têm sido aplicadas para analisar acontecimentos recorrentes foram propostas por Andersen e Gill (AG) [1] e Prentice, Williams e Peterson (PWP) [11].

No que concerne à formulação dos modelos AG e PWP, para o  $i$ -ésimo indivíduo em estudo, as respetivas funções de risco são dadas por

$$h(t; \mathbf{z}_{ik}(t)) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}(t)), \quad t \geq 0, \quad (2)$$

e

$$h(t; \mathbf{z}_{ik}(t)) = h_{0k}(t) \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}(t)), \quad t \geq 0, \quad (3)$$

onde  $h_0(t) \geq 0$  representa a função de risco subjacente comum a todos os acontecimentos,  $h_{0k}(t) \geq 0$  é a função de risco subjacente específica do acontecimento  $k$  e  $\beta = (\beta_1, \dots, \beta_p)'$  é o vetor de parâmetros de regressão. Note-se que estas extensões são na verdade modelos semiparamétricos pois a forma da função de risco subjacente não é especificada.

Em termos gerais, o modelo AG (2) foi proposto para o caso em que os acontecimentos ocorrem de forma ordenada e apresentam igual risco de ocorrerem, pelo que se considera uma função de risco subjacente comum a todos os acontecimentos. Neste modelo, considera-se que todos os indivíduos em estudo contribuem para o conjunto de risco de qualquer acontecimento, seja qual for o número de acontecimentos observados para cada indivíduo. Deste modo, diz-se que o conjunto de risco é não restritivo.

O modelo PWP (3) também surgiu para analisar acontecimentos ordenados, mas pressupõe que o risco de ocorrência de um acontecimento é afetado pela ocorrência do acontecimento que o antecede. Consequentemente, é necessário estratificar os indivíduos segundo a ordem pela qual os acontecimentos ocorrem. Assim, se for possível observar  $k$  acontecimentos, existirão  $k$  estratos ordenados, sendo que a cada um deles estará associada a função de risco subjacente  $h_{0k}(t)$ ,  $k = 1, \dots, K$ . Note-se que neste modelo tanto pode ser obtida uma estimativa global dos parâmetros de regressão  $\beta = (\beta_1, \dots, \beta_p)'$  (alternativa que adotamos de modo a poder ser comparada com a obtida pelo modelo AG), como as estimativas específicas associadas a cada acontecimento  $k$ ,  $\beta_k = (\beta_{k1}, \dots, \beta_{kp})'$ . Relativamente ao conjunto de indivíduos em risco, considera-se que estão em risco para o  $k$ -ésimo acontecimento, apenas os indivíduos aos quais já foi observado o acontecimento  $k - 1$ , o que se traduz num conjunto de risco restritivo.

Em ambos os modelos, a construção do intervalo de risco é feita através dos processos de contagem (*counting process*), em que a escala utilizada se refere ao tempo desde o início do estudo, mas onde os tempos até à ocorrência de cada acontecimento têm como instante inicial o instante em que o acontecimento anterior é observado

(acontecimentos instantâneos) ou em que o acontecimento anterior termina (acontecimentos duradouros). Importa salientar que o modelo PWP permite ainda que o intervalo de risco possa ser formulado segundo o tempo por intervalos (*gap time*), onde a escala de tempo diz respeito ao tempo desde o último acontecimento, sendo que neste caso o relógio reinicia a sua contagem voltando ao instante zero após a ocorrência de cada acontecimento.

Nestes dois modelos semiparamétricos, a estimação do vetor de parâmetros de regressão  $\beta$  é feita pelo método da máxima verosimilhança, pelo que se assume que as observações são independentes. Para isso, é preciso adaptar a função de verosimilhança parcial do modelo de Cox ao contexto dos acontecimentos múltiplos. Todavia, é plausível considerar que os tempos associados ao mesmo indivíduo estejam correlacionados entre si, ou seja, que exista correlação intra-individual. Para ter em conta esse facto, foi desenvolvido um estimador robusto da matriz de covariância – estimador *sandwich* – o qual permite efetuar uma correção na estimativa usual da variância e, por conseguinte, averiguar se as observações estão correlacionadas [9]. Mais detalhes sobre esta temática podem ser encontrados em Kelly e Lim [7].

## 2.2 Modelos paramétricos de Weibull

Conforme citado por Royston e Parmar [14], o sucesso do modelo de Cox fez com que, involuntariamente, seja dedicado pouco esforço ao estudo da função de risco subjacente. Contudo, em determinadas situações, é importante conhecer o comportamento desta função, uma vez que está diretamente relacionada com a forma como a ocorrência dos acontecimentos evolui ao longo do tempo. Além disso, alguns autores como Kwong e Hutton [8] e até o próprio Cox [13], defendem que adotar um modelo paramétrico, quando adequado, pode aumentar a precisão das estimativas dos parâmetros de regressão, o que por sua vez contribui para uma melhor compreensão do fenómeno em estudo. Assim sendo, a adoção de uma abordagem paramétrica pode vir a revelar-se mais apropriada.

Os motivos referidos levaram-nos ao desenvolvimento de dois modelos paramétricos para analisar acontecimentos múltiplos, consoante se considere ou não a estratificação por acontecimento. Tendo por base os modelos de riscos proporcionais e os modelos semiparamétricos AG (2) e PWP (3), a estratégia consistiu em especificar um modelo paramétrico para o tempo. Nesse sentido, optou-se por considerar a distribuição de Weibull com parâmetro de escala  $\lambda > 0$  e parâmetro de forma  $\gamma > 0$ , cuja função de risco pode ser escrita por  $h(t) = \lambda\gamma t^{\gamma-1}$ , para  $t \geq 0$ . A razão pela qual se considerou esta distribuição deve-se ao facto de ocupar um lugar de referência na análise de dados de sobrevivência.

Assim sendo, para o  $i$ -ésimo indivíduo em estudo, as correspondentes funções de risco dos modelos paramétrico Weibull não estratificado (WNE) e paramétrico Weibull estratificado (WE) são definidas por

$$h(t; \mathbf{z}_{ik}(t)) = \lambda\gamma t^{\gamma-1} \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}(t)), \quad t \geq 0, \quad (4)$$

e

$$h(t; \mathbf{z}_{ik}(t)) = \lambda_k \gamma_k t^{\gamma_k-1} \exp(\boldsymbol{\beta}' \mathbf{z}_{ik}(t)), \quad t \geq 0, \quad (5)$$

em que  $\lambda > 0$  e  $\gamma > 0$  denotam os parâmetros de escala e forma comuns a todos os acontecimentos, enquanto  $\lambda_k > 0$  e  $\gamma_k > 0$  representam os parâmetros de escala e forma específicos do acontecimento  $k$ ,  $k = 1, \dots, K$ , respetivamente. Assim, o modelo WNE (4) permite analisar situações em que se assume que o risco de ocorrência dos acontecimentos não se altera. Por outro lado, como no modelo WE (5) são considerados parâmetros de escala e forma específicos de cada acontecimento, este modelo abrange situações em que se admite que os acontecimentos têm riscos de ocorrência diferentes.

Nestes dois modelos paramétricos, os métodos de inferência estatística também são baseados na teoria assintótica de máxima verosimilhança. Para estimar os vários parâmetros de cada modelo, admite-se que as observações são censuradas à direita e que os tempos até cada acontecimento e os tempos de censura são independentes. Deste modo, assumindo que a censura é não informativa, para

ambos os modelos a expressão geral da função de verosimilhança é dada por

$$L = \prod_{i=1}^n \prod_{k=1}^K h(t_{ik}; \mathbf{z}_{ik}(t_{ik}))^{\delta_{ik}} S(t_{ik}; \mathbf{z}_{ik}(t_{ik})), \quad (6)$$

onde  $t_{ik}$  denota o tempo de observação do indivíduo  $i$  referente ao  $k$ -ésimo acontecimento,  $\delta_{ik}$  é a variável que caracteriza o estado do indivíduo em relação ao acontecimento  $k$  e  $h(t_{ik}; \mathbf{z}_{ik}(t_{ik}))$  e  $S(t_{ik}; \mathbf{z}_{ik}(t_{ik}))$  representam a função de risco e a função de sobrevivência associadas ao modelo que estiver a ser ajustado, respetivamente. Embora a construção das funções de verosimilhança dos modelos WNE (4) e WE (5) tenha por base a mesma expressão (6), a função resultante difere consoante a situação, dado que no segundo modelo existe estratificação.

### 3 Uma aplicação com dados simulados

Com o propósito de exemplificar a aplicação dos quatro modelos formulados anteriormente, recorreu-se ao *software* estatístico R, versão 3.5.0. Embora existam vários conjuntos de dados reais disponíveis neste *software* (por exemplo, em [15]), optou-se por realçar a possibilidade de este poder ser utilizado para simular acontecimentos recorrentes, em particular através do *package* *survsim* [10].

Por conseguinte, foi gerada uma amostra constituída por  $n = 1000$  indivíduos, onde se definiu que os mesmos podiam sofrer no máximo  $K = 8$  acontecimentos, de forma a ter um número razoável de indivíduos nos estratos respeitantes aos últimos acontecimentos e, assim, evitar que sejam obtidas estimativas muito instáveis. Relativamente ao tempo de *follow-up*, estipulou-se que este seria igual a 1825 dias (o equivalente a 5 anos) pois, em geral, é necessário que o tempo seja suficientemente longo para que o acontecimento possa repetir-se.

Uma vez que o *package survsim* apenas permite que os tempos até cada acontecimento sejam gerados por intermédio das distribuições de Weibull, log-normal ou log-logística, optou-se pela distribuição de Weibull. Para os tempos de censura, acresce ainda a possibilidade de considerar a distribuição uniforme, algo que é usual neste contexto. No entanto, optou-se por escolher a mesma distribuição para ambos os casos. Assim sendo, definiu-se que os tempos seguem distribuições de Weibull com parâmetros de forma distintos, mas maiores do que 1 de modo a que os acontecimentos tivessem riscos de ocorrência diferentes e que as correspondentes funções de risco fossem crescentes.

Por último, o *package survsim* permite que as covariáveis sejam geradas através de três distribuições distintas, tendo-se decidido explorar as três possibilidades. Simulou-se então uma variável discreta com distribuição de Bernoulli com probabilidade de sucesso igual a 0.5, correspondente à covariável  $x$ ; uma variável contínua com distribuição uniforme no intervalo  $[0, 1]$ , correspondente à covariável  $x.1$ ; e uma outra variável contínua com distribuição gaussiana padrão, correspondente à covariável  $x.2$ . Nesta etapa, estipulou-se que o efeito de cada covariável seria o mesmo em todos os acontecimentos, sendo que os valores fixados foram  $\beta_x = 0.6$ ,  $\beta_{x.1} = 0.03$  e  $\beta_{x.2} = 0.75$ , respetivamente. Importa salientar que o propósito deste trabalho não é efetuar um estudo de simulação, mas sim utilizar a simulação como um meio de obter um conjunto de dados adequado à aplicação dos dois modelos propostos. Assim, a partir deste momento o conjunto de dados simulados será tratado como se fossem dados reais.

Na Tabela 1, resumiu-se a informação relevante sobre a constituição e evolução do conjunto de indivíduos em risco por acontecimento. Deste modo, obtém-se uma visão global sobre as características dos dados que, posteriormente, pode contribuir para uma melhor compreensão dos resultados obtidos no ajustamento dos modelos. Procedeu-se então à implementação dos modelos, em que para os dois modelos semiparamétricos (AG e PWP) utilizou-se o *package survival* [15] e para os dois modelos paramétricos (WNE e

Tabela 1: Resumo da informação sobre os dados simulados.

	Número do acontecimento							
	1	2	3	4	5	6	7	8
Conjunto de risco	1000	355	143	83	52	27	17	11
Acontecimentos observados	355	143	83	52	27	17	11	8
Percentagem de censura (%)	64.5	59.7	42.0	37.3	48.1	37.0	35.3	27.3

WE) recorreu-se ao *package* `straweb` [4]. Os resultados obtidos no ajustamento de cada modelo encontram-se compilados na Tabela 2. Como é possível constatar, em todos os modelos as covariáveis  $\mathbf{x}$  e  $\mathbf{x}.2$  têm influência significativa sobre o tempo até a ocorrência do acontecimento e a covariável  $\mathbf{x}.1$  não. Ao comparar os valores do parâmetro associado a cada covariável definido no processo de simulação com as respetivas estimativas, observa-se que os modelos PWP e WE são os que melhor conseguem estimar os valores fixados. Na verdade, o modelo PWP é o que apresenta melhores resultados em relação ao efeito das covariáveis  $\mathbf{x}$  e  $\mathbf{x}.1$ , enquanto o modelo WE se destaca em relação ao efeito da covariável  $\mathbf{x}.2$ .

Considerando a estimativa usual do erro padrão ( $EP(\hat{\beta}_j)$ ), verifica-se que estas são praticamente iguais entre modelos não estratificados (AG e WNE) e modelos estratificados (PWP e WE). Nos modelos semiparamétricos pode-se comparar ainda a estimativa usual do erro padrão com a respetiva estimativa robusta, onde se observa que  $EP_r(\hat{\beta}_j)$  é ligeiramente superior a  $EP(\hat{\beta}_j)$ . Segundo Kelly e Lim [7], este facto está de acordo com a potencial existência de correlação intra-individual.

Em ambas as abordagens, semiparamétrica e paramétrica, o modelo mais adequado para estes dados foi o que considerou a estratificação, o que é coerente com o facto de os dados terem sido simulados de modo a que os acontecimentos tivessem riscos de ocorrência distintos. De facto, através do critério de informação de Akaike (*AIC*), na abordagem semiparamétrica, verifica-se que o modelo

Tabela 2: Estimativas dos parâmetros de regressão para cada modelo.

Covariável/Modelo	$\hat{\beta}_j$	$\exp(\hat{\beta}_j)$	$EP(\hat{\beta}_j)$	$EP_r(\hat{\beta}_j)$	Valor- $p$
x					
AG	0.806	2.238	0.081	0.107	5.84e-14
WNE	0.829	2.290	0.082	—	< 2e-16
PWP	0.651	1.917	0.083	0.088	1.20e-13
WE	0.683	1.981	0.083	—	2.22e-16
x.1					
AG	0.064	1.066	0.131	0.173	0.714
WNE	0.063	1.065	0.133	—	0.636
PWP	0.048	1.049	0.133	0.137	0.725
WE	0.057	1.059	0.135	—	0.671
x.2					
AG	0.861	2.365	0.046	0.055	< 2e-16
WNE	0.871	2.389	0.045	—	< 2e-16
PWP	0.720	2.055	0.048	0.047	< 2e-16
WE	0.765	2.150	0.048	—	< 2e-16

PWP apresenta um valor  $AIC$  inferior ao obtido para o modelo AG ( $AIC_{PWP} = 7133.034$  e  $AIC_{AG} = 9035.034$ ) e, na abordagem paramétrica, observa-se que o valor  $AIC$  mais baixo encontra-se associado ao modelo WE ( $AIC_{WE} = 10825.300$  e  $AIC_{WNE} = 11015.160$ )<sup>5</sup>.

Na abordagem paramétrica, pode-se ainda avaliar a adequabilidade do modelo, de forma informal, representando graficamente as estimativas da função de sobrevivência obtidas pelo estimador de Kaplan-Meier e pelo modelo nulo<sup>6</sup> que estiver a ser considerado, tendo-se obtido o gráfico apresentado à esquerda na Figura 1. Através deste gráfico, verifica-se que as estimativas baseadas no modelo de Weibull são bastante próximas das estimativas de Kaplan-Meier,

<sup>5</sup>Note-se que a comparação dos valores  $AIC$  obtidos para cada modelo só pode ser feita dentro de cada abordagem, visto que o tipo de verosimilhança considerada não é o mesmo.

<sup>6</sup>O modelo é ajustado sem incluir as covariáveis.

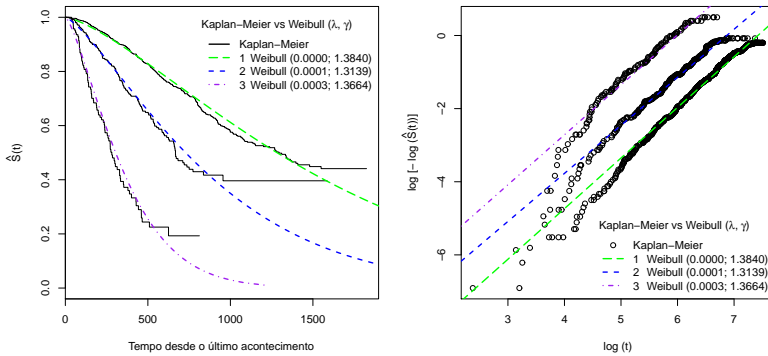


Figura 1: Estimativas da função de sobrevivência de Kaplan-Meier e para a distribuição de Weibull, referentes aos três primeiros acontecimentos.

excepto na parte final. Relembre-se que os dados foram propositalmente simulados através da distribuição de Weibull, pelo que este resultado já seria de esperar. Outra forma de realizar esta análise, consiste em considerar a representação de  $\log[-\log(\hat{S}(t))]$  versus  $\log(t)$  para cada acontecimento, originando o gráfico apresentado à direita. Analisando este gráfico, confirma-se que de facto assumem uma forma razoavelmente linear, embora tal não aconteça em alguns dos menores tempos observados. Note-se que, para não sobrecarregar o gráfico, decidiu-se apenas representar as estimativas referentes aos três primeiros acontecimentos/estratos.

## 4 Conclusões e trabalho futuro

De um modo geral, os dois modelos paramétricos baseados na distribuição de Weibull (WNE e WE), revelaram-se uma alternativa bastante adequada às duas extensões do modelo semiparamétrico de Cox que foram propostas para analisar acontecimentos recorrentes,

os modelos AG e PWP.

A implementação dos dois novos modelos foi efetuada com recurso ao *package straweib* [4], originalmente criado para lidar com a violação do pressuposto de riscos proporcionais, permitindo uma função de risco subjacente específica para cada estrato. Assim, o que se fez foi adaptar o código deste *package* ao caso em que são observados vários acontecimentos por indivíduo, utilizando a variável que indica o número do acontecimento como variável de estratificação.

Quando a função de risco subjacente é corretamente especificada, sabe-se que os modelos paramétricos evidenciam maior eficiência comparativamente aos modelos semiparamétricos. Embora a aplicação dos modelos WNE e WE a este conjunto de dados simulados, em particular, não tenha melhorado a precisão das estimativas dos parâmetros de regressão, estes modelos têm a vantagem de permitir estimar de forma suave a função de risco. Além do mais, conhecer a distribuição associada ao tempo até cada acontecimento contribui para uma melhor compreensão sobre o modo como a multiplicidade de acontecimentos evolui ao longo do tempo.

A abordagem aplicada neste trabalho consistiu em considerar a distribuição de Weibull para especificar a forma da função de risco subjacente e, consequentemente, obter um modelo totalmente paramétrico. Porém, consoante o caso de estudo, tem interesse ponderar outras distribuições que podem vir a revelar-se mais apropriadas ou até mesmo ter maior flexibilidade para captar a forma como o risco evolui ao longo do tempo. Nesse sentido, recomenda-se explorar o *package flexsurv* [6] para implementar esses novos modelos.

Apesar de esta investigação ter sido direcionada para acontecimentos recorrentes, isso não significa que estes dois modelos paramétricos não possam ser utilizados para analisar acontecimentos de tipos diferentes. Uma outra temática a explorar futuramente seria desenvolver um estimador robusto para a matriz de covariância, à semelhança daquele que foi desenvolvido para as extensões do modelo de Cox.

## Agradecimentos

O primeiro autor agradece à Sociedade Portuguesa de Estatística (SPE), pela bolsa que lhe foi concedida para participar no “XXIII Congresso da SPE”. Esta investigação foi parcialmente financiada por Fundos Nacionais, através da Fundação para a Ciência e a Tecnologia (FCT), no âmbito dos projetos UID/MAT/00006/2013 (Centro de Estatística e Aplicações) e UID/MAT/04674/2013 (Centro de Investigação em Matemática e Aplicações).

## Referências

- [1] Andersen, P. K. e Gill, R. D. (1982). Cox’s regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4), 1100–1120.
- [2] Cox, D. R. (1972) Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34(2), 187–220.
- [3] Ferreira, I. M. S. (2016). *Modelos para acontecimentos múltiplos*. (Dissertação de Mestrado). Universidade da Madeira, Funchal, Portugal.
- [4] Gu, X. e Balasubramanian, R. (2013). straweib: Stratified Weibull Regression Model. *Package do R versão 1.0*. URL: <https://github.com/cran/straweib>.
- [5] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag, New York.
- [6] Jackson, C., Metcalfe, P. e Amdahl, J. (2017). flexsurv: Flexible Parametric Survival and Multi-State Models. *Package do R versão 1.1*. URL: <https://CRAN.R-project.org/package=flexsurv>.
- [7] Kelly, P. J. e Lim, L. L-Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, 19(1), 13–33.
- [8] Kwong, G. P. S. e Hutton, J. L. (2003). Choice of parametric models in survival analysis: applications to monotherapy for epilepsy and cerebral palsy. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(2), 153–168.

- [9] Lin, D. Y. e Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84(408), 1074–1078.
- [10] Moriña, D. e Navarro, A. (2015). *survsim: Simulation of Simple and Complex Survival Data*. Package do R versão 1.1.5. URL: <https://CRAN.R-project.org/package=survsim>.
- [11] Prentice, R. L., Williams, B. J. e Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2), 373–379.
- [12] R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [13] Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, 9(3), 439–455.
- [14] Royston, P. e Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15), 2175–2197.
- [15] Therneau, T. M. (2015). *survival: A package for survival analysis in S*. Package do R versão 2.41-3. URL: <https://CRAN.R-project.org/package=survival>.
- [16] Therneau, T. M. e Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.