



## O uso modelos de fração de cura discreto na análise de resultados oncológicos no tratamento de sarcomas pélvicos

André F. B. Menezes<sup>1</sup>, Ricardo Puziol<sup>2</sup> e Josmar Mazucheli<sup>1</sup>

<sup>1</sup>Universidade Estadual de Maringá, Departamento de Estatística, DEs, PR, Brasil.

<sup>2</sup>Medical School, University of São Paulo, Ribeirão Preto, SP, Brasil

### Resumo

Modelos de fração de cura são amplamente utilizados quando existem indivíduos não-suscetíveis ao evento de interesse. Uma vez que há presença desses indivíduos, independente da natureza do tempo observado, em geral, assume-se uma distribuição contínua para realizar a modelagem. Todavia, em muitas situações a medida de tempo é discreta como em estudos de câncer em que o tempo é medido em número de dias, meses ou anos. Nos últimos anos, para se evitar o uso de distribuições contínuas na análise de dados discretos, diversas distribuições discretas foram propostas, entretanto poucas ou quase nenhuma foram utilizadas em modelagens com presença de fração de cura. No presente trabalho, ilustramos uma nova abordagem usando modelos discretos com fração de cura em um estudo realizado por Puchner et al. (2017) a respeito do tratamento de sarcomas pélvicos em pacientes do sexo feminino e masculino. Os resultados mostraram que os modelos de fração de cura devem ser considerados para uma melhor análise por oferecerem uma precisão melhor. Além disso, os resultados inferenciais mostraram uma boa performance com baixo custo computacional o que nos leva a conclusão de que o uso de modelos discretos pode ser uma boa alternativa para a análise de dados de sobrevivência com fração de cura.

**Palavras chave:** análise de sobrevivência; fração de cura; distribuições discretas; modelo de mistura.

## 1 Introdução

Em geral, a modelagem e análise de dados relacionados a fração de pacientes que nunca experimentam o evento de interesse desempenham um papel importante em pesquisas médicas, especialmente em pesquisas oncológicas (HSU; TODOEM; KIM, 2016). Todavia, as técnicas usuais de análise de sobrevivência, como o modelo de riscos proporcionais de Cox (COX, 1972), não permitem uma estimativa direta da fração de cura, surge a necessidade de introduzir novos modelos paramétricos que sejam suficientes para permitir uma estimativa da fração de cura. Diferentes abordagens apresentadas na literatura para modelar a fração de cura, especialmente para dados de vida univariados, podem ser vistas em: Angelis et al. (1999); Yin e Ibrahim (2005); Lambert et al. (2006); Achcar, Coelho-Barros e Mazucheli (2012); Othus et al. (2012), dentre muitos outros.

No presente estudo, será ilustrado uma nova abordagem usando modelos discretos de fração de cura em um estudo realizado por Puchner et al. (2017) a respeito do tratamento de sarcomas pélvicos. É importante destacar que os sarcomas pélvicos são cirurgicamente difíceis de serem removidos devido à proximidade anatômica da pelvis à muitas estruturas neurovasculares, tratos urinários e intestinais. Por consequência, isto gera maus resultados oncológico e alta taxa de complicações. O conjunto de dados é provenientes de um estudo de coorte retrospectivo, em que foi considerado 147 pacientes com sarcoma pélvico entre 1980 e 2012. O evento de interesse é o óbito do paciente, logo a variável resposta  $T$  é o tempo, em meses completos, desde o início do tratamento até o óbito do paciente, censura ou o fim do estudo.

## 2 Metodologia

Seja  $T$  uma variável aleatória representando o tempo até a ocorrência de um determinado evento de interesse. A função de sobrevivência, definida como  $S(t) = \Pr(T \geq t)$ , indica a probabilidade de um indivíduo sobreviver mais do que um certo tempo  $t$ . Usualmente, assume-se que  $S(t)$  converge para zero a medida  $t$  tende para o infinito, ou seja, todos os indivíduos estão sujeitos a ocorrência do evento. No entanto, essa suposição pode não condizer com a realidade quando, por exemplo, existem indivíduos imunes ou curados na população.

Considerando os dados descritos na introdução, onde fica evidente a presença de indivíduos imunes, para descrever o tempo até a morte por sarcoma foi empregado o modelo de mistura proposto por Berkson e Gage (1952), no qual inclui explicitamente um parâmetro que descreve a taxa de curados. Seguindo notação de Farewell (1982) assume-se que a função de sobrevivência é expressa por

$$S(t) = \pi + (1 - \pi) S_0(t), \quad (1)$$

em que  $\pi \in (0, 1)$  é a proporção de imunes e  $S_0(t)$  é a função de sobrevivência dos indivíduos suscetíveis.

Sob o ponto de vista paramétrico existem diversas escolhas para  $S_0(t)$ . É importante considerar modelos que representam a natureza da variável observada. No presente estudo, uma vez que o tempo de sobrevida foi medido em escala discreta, meses completos, serão consideradas distribuições discretas.

De acordo com Nakagawa e Osaki (1975) a versão discreta de uma variável aleatória contínua  $X$ , com função de sobrevivência  $S_X(x) = \Pr(X \geq x)$ , tem função massa de probabilidade definida por

$$\Pr(T = k) = S_X(k) - S_X(k - 1), \quad k = 0, 1, 2, \dots \quad (2)$$

em que  $T = \lfloor X \rfloor$  indica o maior inteiro menor ou igual a  $X$ . Uma propriedade interessante deste método é que mesmo preserva a expressão da função de sobrevivência de  $X$ .

Com exceção das distribuições Geométrica e Weibull discreta (NAKAGAWA; OSAKI, 1975), os outros modelos adotados neste trabalho foram recentemente propostos na literatura utilizando (2). De fato, a versão discreta da distribuição Gamma, Weighted Lindley, Half-Normal Generalizada foram proposta, respectivamente, por Chakraborty e Chakravarty (2012), Oliveira (2016) e Cardoso e Mazucheli (2017).

Ressalta-se que a inferência para os parâmetros dos modelos adotados foi realizada sob o paradigma da verossimilhança. Sem perdas de generalidades, considerando uma amostra aleatória  $(t_i, \delta_i), i = 1, \dots, n$ , a função log-verossimilhança do modelo (1) pode ser escrita como

$$\ell = \sum_{i=1}^n \delta_i \log(1 - \pi) + \sum_{i=1}^n \delta_i \log f_0(t_i) + \sum_{i=1}^n (1 - \delta_i) \log [\pi + (1 - \pi) S_0(t_i)] \quad (3)$$

em que  $\delta_i$  é uma variável indicadora de censura, isto é,  $\delta_i = 1$  para tempo observados e  $\delta_i = 0$  para tempos censurados e  $f_0(t_i)$  é a função de probabilidade para dos indivíduos suscetíveis.

As função de log-verossimilhança dos modelos considerados foram implementadas no software R (R Core Team, 2016) e as estimativas foram obtidas via maximização numérica utilizando a função `optim`. Os erros padrão foram calculados por meio da matriz de informação observada de Fisher.

## 3 Resultados e Discussões

Na Figura 1, são apresentadas as curvas de Kaplan-Meier versus os modelos ajustados tanto na presença como ausência da fração de cura de acordo com o sexo. Observando as curvas de Kaplan-Meier versus os modelos ajustados sem considerar a presença da fração de cura, podemos observar que todos os modelos ilustram um ajuste precário o que nos leva a concluir que o uso de modelos de fração de cura devem ser considerados para uma melhor análise.

Admitindo-se então a modelagem com fração de cura, dos resultados da Tabela 1, pode-se concluir que todos os modelos, em termos computacionais, não mostraram instabilidade computacional e os parâmetros convergiram com sucesso. Além disso, para comparar os modelos propostos, é também apresentado na Tabela 1 as estimativas do AICc e BIC. Observando conjuntamente os valores de AICc e BIC pode-se concluir que as distribuições Geométrica e Weighted Lindley na presença de fração de cura obtiveram os melhores ajustes para ambos os sexos. Por outro lado, olhando para as curvas de Kaplan-Meier (veja a Figura 1) na presença da covariável sexo, todos os modelos propostos capturam a fração de cura com boa precisão. Ademais, a proporção estimada de pacientes mulheres não-suscetíveis ao óbito é de 39,68%, já a proporção estimada de pacientes homens não-suscetíveis ao óbito foi de 35,01%.

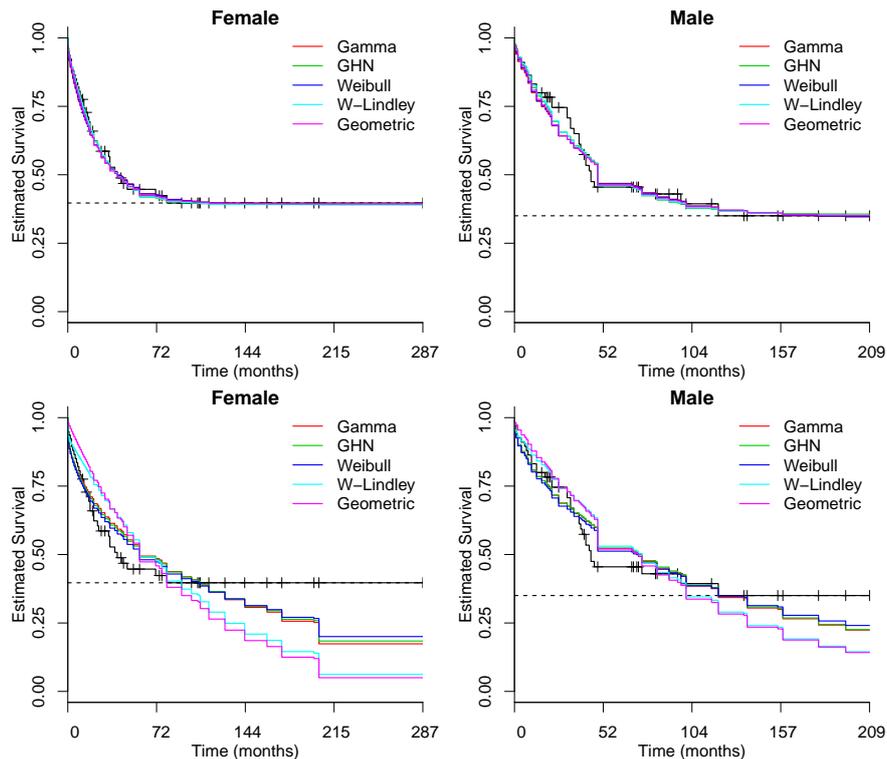


Figura 1: Curvas de sobrevivência estimadas de acordo com o gênero. (painel superior: modelos com fração de cura; painel inferior: modelos sem fração de cura.)

Portanto, dado que a natureza dos dados é discreta e os resultados inferenciais mostram uma boa performance e precisão, concluímos que há um bom indicativo para o uso de modelos discretos como alternativa a análise de dados de sobrevivência.

## Referências

- ACHCAR, J.; COELHO-BARROS, E.; MAZUCHELI, J. Cure fraction models using mixture and non-mixture models. *Tatra Mountains Mathematical Publications*, v. 51, n. 1, p. 1–9, 2012.
- ANGELIS, R. D. et al. Mixture models for cancer survival analysis: Application to population-based data with covariates. *Statistics in Medicine*, John Wiley & Sons, Ltd., v. 18, n. 4, p. 441–454, 1999.
- BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, Taylor & Francis, v. 47, n. 259, p. 501–515, 1952.
- CARDOSO, J. C.; MAZUCHELI, J. A distribuição Half-Normal generalizada discreta. *Ciência & Natura*, 2017. Submetido.
- CHAKRABORTY, S.; CHAKRAVARTY, D. Discrete Gamma distributions: Properties and parameter estimations. *Communications in Statistics - Theory and Methods*, Taylor & Francis, v. 41, n. 18, p. 3301–3324, 2012.
- COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 34, n. 2, p. 187–220, 1972.
- FAREWELL, V. T. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, v. 38, n. 4, p. 1041–1046, 1982.
- HSU, W.-W.; TODEM, D.; KIM, K. A sup-score test for the cure fraction in mixture models for long-term survivors. *Biometrics*, Wiley Online Library, v. 72, n. 4, p. 1348–1357, 2016.
- LAMBERT, P. C. et al. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, v. 8, n. 3, p. 576–594, 2006.

NAKAGAWA, T.; OSAKI, S. The discrete Weibull distribution. *IEEE Transactions on Reliability*, n. 5, p. 300–301, 1975.

OLIVEIRA, R. P. *Aplicações de Alguns Métodos de Discretização nas Distribuições Lindley Contínuas*. Dissertação (Mestrado) — Universidade Estadual de Maringá, 2016.

OTHUS, M. et al. Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research*, v. 18, n. 14, p. 3731–3736, 2012.

PUCHNER, S. E. et al. Oncological and surgical outcome after treatment of pelvic sarcomas. *PLoS ONE*, v. 12, n. 2, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <https://www.R-project.org/>.

YIN, G.; IBRAHIM, J. G. Cure rate models: A unified approach. *Canadian Journal of Statistics*, Wiley-Blackwell, v. 33, n. 4, p. 559–570, 2005.

Tabela 1: Resumo dos modelos ajustados.

Sex	Model	Parameter	Estimates (S.E.)	95% C.I.	AICc	BIC
Female	Gamma	$\alpha$	0.8629 (0.1883)	(0.4937, 1.2320)	376.9486	383.7369
		$\theta$	29.0278 (10.4531)	(8.5401, 49.5155)		
		$\pi$	0.3916 (0.0702)	(0.2540, 0.5292)		
	GHN	$\alpha$	0.7614 (0.1130)	(0.5398, 0.9829)	374.9886	381.7769
		$\theta$	28.1417 (5.2376)	(17.8761, 38.4073)		
		$\pi$	0.3924 (0.0682)	(0.2587, 0.5260)		
	Weibull	$\beta$	0.9320 (0.1436)	(0.6505, 1.2136)	377.2044	383.9927
		$q$	0.9500 (0.0233)	(0.9044, 0.9957)		
		$\pi$	0.3913 (0.0703)	(0.2534, 0.5291)		
	W-Lindley	$\alpha$	0.2883 (0.1170)	(0.0591, 0.5176)	374.1546	380.9429
		$\theta$	0.0482 (0.0114)	(0.0258, 0.0706)		
		$\pi$	0.3907 (0.0686)	(0.2563, 0.5252)		
Geom	$p$	0.0407 (0.0078)	(0.0254, 0.0560)	375.2614	379.8424	
	$\pi$	0.3938 (0.0689)	(0.2588, 0.5289)			
Male	Gamma	$\alpha$	0.9020 (0.2082)	(0.4939, 1.3102)	380.8311	387.1146
		$\theta$	46.9955 (20.3795)	(7.0523, 86.9387)		
		$\pi$	0.3402 (0.0837)	(0.1761, 0.5043)		
	GHN	$\alpha$	0.8110 (0.1271)	(0.5618, 1.0601)	379.3692	385.6528
		$\theta$	45.6407 (9.3719)	(27.2721, 64.0093)		
		$\pi$	0.3561 (0.0759)	(0.2073, 0.5048)		
	Weibull	$\beta$	0.9611 (0.1660)	(0.6358, 1.2864)	380.9814	387.2649
		$q$	0.9723 (0.0158)	(0.9414, 1.0033)		
		$\pi$	0.3420 (0.0847)	(0.1759, 0.5081)		
	W-Lindley	$\alpha$	0.2846 (0.1320)	(0.0259, 0.5434)	379.2707	385.5542
		$\theta$	0.0305 (0.0083)	(0.0143, 0.0468)		
		$\pi$	0.3502 (0.0769)	(0.1995, 0.5008)		
Geom	$p$	0.0244 (0.0056)	(0.0134, 0.0355)	378.8452	383.0996	
	$\pi$	0.3475 (0.0787)	(0.1932, 0.5018)			