Modelo de Regressão Kumaraswamy-Weibull com Fração de Cura em Análise de Sobrevivência

Beatriz Leal Simões e Silva *
Universidade de Brasília

Gustavo Pompeu da Silva †
Universidade de Brasília

Juliana Betini Fachini-Gomes ‡
Universidade de Brasília

4 de dezembro de 2018

1 Introdução

Uma importante área de Estatística é a Análise de Sobrevivência que tem como objetivo analisar tempos até o acontecimento de um determinado evento, e juntamente com a censura. O conjunto de dados então é formado por amostras que incluem o tempo, censura e demais covariáveis que influenciam o estudo em questão.

O trabalho é composto pelo desenvolvimento de um modelo de regressão baseado na distribuição Kumaraswamy, utilizando a função Weibull para compor a função de distribuição Kumaraswamy-Weibull devido ao fato da Weibull ser uma distribuição com muitas aplicações nessa área de estudo. Além disso, foi utilizado o conceito de fração de cura, que acontece quando é assumido que parte da população foi curada durante o estudo, e os conceitos de inserção de covariáveis em um dos parâmetros da distribuição de probabilidade utilizada.

2 Material e Métodos

2.1 Material

O conjunto de dados são provenientes de uma coorte com 862 pacientes com câncer internados na UTI do Inca, publicado por Soares e cols. (2006), cujo objetivo principal do estudo foi avaliar fatores associados à sobrevivência.

As covariáveis incluídas no estudo são: sexo do paciente; idade em anos completos; tipo de tumor, categorizado em sólido localizado, metastático e hematológico; desnutriçãodo paciente, perda de peso recente acima de $> 10\%$ ou $IMC < 18$; presença de comorbidades severas; presença de leucopenia

2.2 Métodos

Para construir um modelo de regressão na Análise de Sobrevivência precisa-se entender como o tempo se afeta pelas covariáveis incluídas no modelo. o estimador de Kaplan-Meier é uma técnica utilizada para descrever o tempo em relação as demais variáveis. O gráfico da curva de sobrevivência de Kaplan-Meier para os dados em questão não tende à zero ao final da curva, o que indica a utilização de um modelo de fração de cura.

*Address: UnB, Universidade de Brasília, Brasília, Brasil. E-mail: beatrizleal@gmail.com
†Address: UnB, Universidade de Brasília, Brasília, Brasil. E-mail: gustavopompeu_@hotmail.com
‡Address: UnB, Universidade de Brasília, Brasília, Brasil. E-mail: jfachini@unb.br

Nenhum dos modelos paramétricos usuais tiveram ajuste aceitável nesse conjunto de dados. Logo, o melhor caminho foi buscar um meio de ajustar por outro modelo, através da distribuição Kumaraswamy-Weibull com fração de cura.

A distribuição Kumaraswamy generalizada tem como função de distribuição acumulada:

$$F(t; a, b) = 1 - [1 - G(t)^a]^b \tag{1}$$

E função densidade de probabilidade:

$$f(t; a, b) = abg(t)G(t)^{a-1}[1 - G(t)^a]^{b-1} \tag{2}$$

onde $a > 0$, $b > 0$ e $G(t)$ é uma f.d.a arbitrária que pode ser qualquer distribuição de probabilidades, e $g(t)$ é sua f.d.p.

A distribuição Weibull foi escolhida por seu frequente uso em Análise de Sobrevivência. $G(t)$ será a f.d.a da Weibull, que é definida por:

$$G(t; \lambda, c) = 1 - exp[-(\lambda t)^c] \tag{3}$$

em que $c > 0$ é o parâmetro de forma e $\lambda > 0$ é o parâmetro de escala.

Então, a f.d.p da distribuição K-Weibull será:

$$f(t; \lambda, c, a, b) = abc\lambda^c t^{c-1} exp[-(\lambda t)^c]\{1 - exp[-(\lambda t)^c]\}^{a-1}\{1 - \{1 - exp[-(\lambda t)^c]\}^a\}^{b-1} \tag{4}$$

E a f.d.a:

$$F(t; \lambda, c, a, b) = 1 - \{1 - \{1 - exp[-(\lambda t)^c]\}^a\}^b \tag{5}$$

E como temos a relação $S(t) = 1 - F(t)$, a função de sobrevivência da distribuição K-Weibull é dada por:

$$S(t; \lambda, c, a, b) = \{1 - \{1 - exp[-(\lambda t)^c]\}^a\}^b \tag{6}$$

Para se obter um modelo de regressão da distribuição K-Weibull, vamos fazer uma reparametrização de dois parâmetros: $\lambda = exp\{-\mu\}$ e $c = \frac{1}{\sigma}$, onde $-\infty < \mu < \infty$, e $\sigma > 0$.

Ao considerar o vetor de variáveis regressoras $\mathbf{x} = (x_0, x_1, ...x_p)^T$, consideramos $\mu = \mathbf{x}^T\boldsymbol{\beta}$, em que

$\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T$ é o vetor de parâmetros desconhecidos associados às covariáveis. Então, nas funções da K-Weibull teremos:

$$\lambda = exp\{-(\beta_0 x_0 + \beta_1 x_1 + ... + \beta_p x_p)\} \tag{7}$$

Um modelo de fração de cura tem como função de sobrevivência:

$$S_{pop}(t) = \phi + (1 - \phi)S_{mod}(t) \tag{8}$$

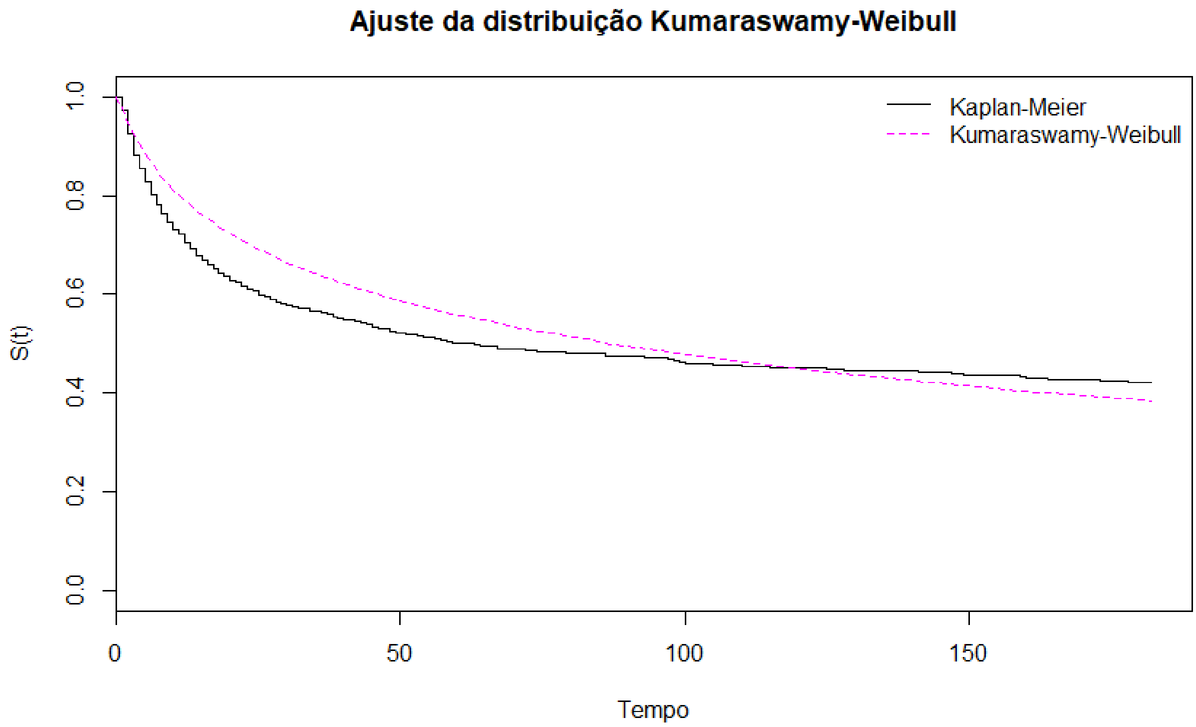
e função densidade de probabilidade:

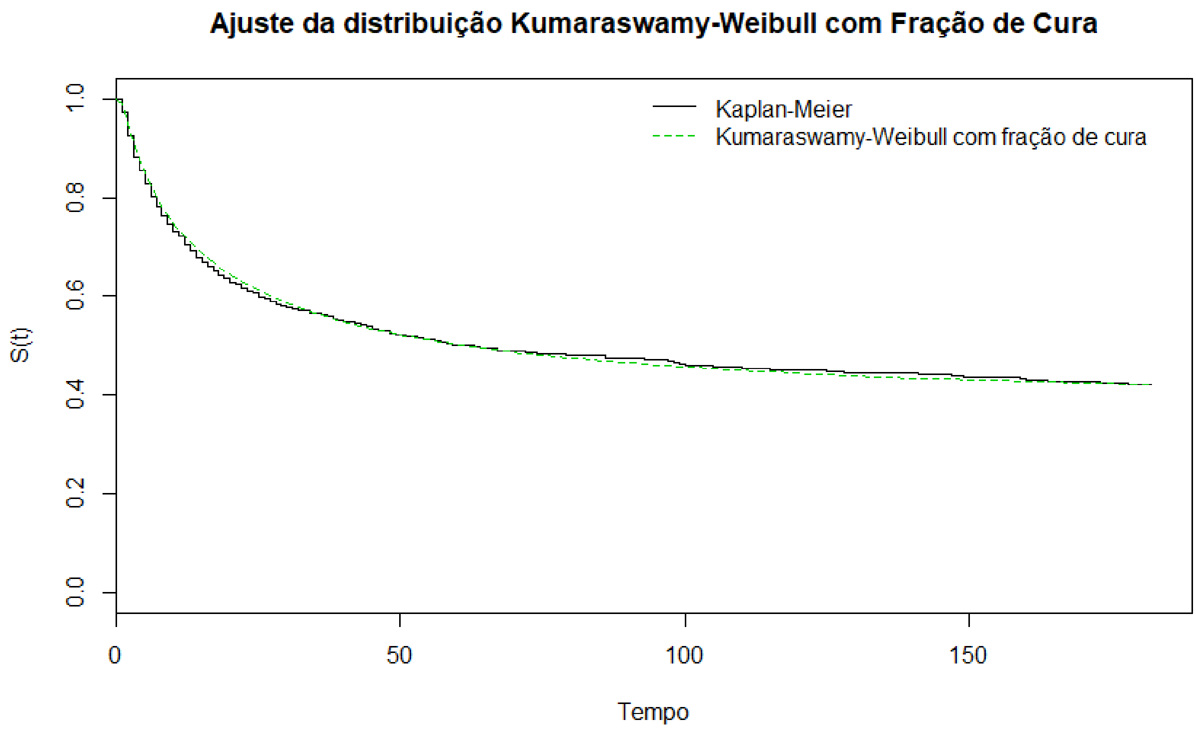
$$f_{pop}(t) = (1 - \phi)f_{mod}(t) \tag{9}$$

Onde $S_{mod}(t)$ e $f_{mod}(t)$ são a função de sobrevivência e a f.d.p de um modelo paramétrico qualquer, respectivamente, e que no nosso caso são as funções do modelo K-Weibull. ϕ é um parâmetro que representa a fração da população que foi curada durante o estudo, ou seja, que ao final do estudo não estava mais sob risco de falha.

3 Resultados

O ajuste dos modelo Kumaraswamy-Weibull, sem, e com fração de cura, respectivamente, se deram por:





Ambos os ajustes acima são no modelo sem covariáveis.

Para introduzir as covariáveis, analisamos através das curvas de sobrevivência de Kaplan-Meier por covariável categórica, e dos testes de Log-Rank e de Wilcoxon, foram definidas as covariáveis a serem incorporadas no modelo de regressão. E através do processo Forward, foram incluídas apenas as variáveis significativas. Todas as covariáveis, com excessão da idade, são categóricas, então foi necessário a criação de variáveis *dummy* para introduzir as informações qualitativas corretamente.

Os parâmetros do modelo final foram:

Variável	Estimativa	Erro Padrão	p-valor
β_0	-1.96271266	0.35214718	0.00000
β_1	-0.75790481	0.26003796	0.00356
β_2	-0.49832606	0.22399431	0.02610
β_3	-0.73964719	0.20828554	0.00038
β_4	-0.01288992	0.00473482	0.00648
β_5	-0.68897000	0.26286074	0.00877
σ	2.73859993	0.11105442	-
a	47.62133747	15.45617199	-
b	0.13761553	0.02949474	-
ϕ	0.33645631	0.03225521	-

Onde:

- β_1 é referente à variável de leucopenia (X_1)
- β_2 e β_3 são referentes à variável do tipo de tumor do paciente (X_2 e X_3)
- β_4 é referente à variável da idade do paciente (X_4)
- β_5 é referente à variável de desnutrição do paciente (X_5)

A presença de um tumor Hematológico ou Metastático diminui a probabilidade de sobrevivência do paciente, quando comparado com os pacientes que possuem tumor Sólido Localizado. Tal como, a presença de leucopenia também diminui a probabilidde de sobrevivência do paciente. Quanto mais velho for o paciente, menor as chances de sobrevivência do mesmo.

Em relação à presença de sinais de desnutrição (perda de peso recente acima de 10% ou IMC <18) diminui a probabilidade de sobrevivência do paciente.Por fim, a presença de comorbidades severas não influencia significativamente na probabilidade de sobrevivência dos pacientes, assim como o sexo do paciente.

4 Conclusão

Esse trabalho buscou desenvolver de forma computacional um modelo de regressão aplicando uma função de distribuição não trivial em Análise de Sobrevivência.

Aplicado a fração de cura no modelo de regressão proposto com base na distribuição Kumaraswamy-Weibull obteve um ajuste extremamente próximo a curva de sobrevivência para o banco de dados utilizado, o modelo foi preciso. Os resultados apresentados são compatíveis com a modelagem proposta e é bem explicado pela função considerada.

5 Bibliografia

FACHINI, Juliana Betini. **Modelos de regressão com e sem fração de cura para dados bivariados em análise de sobrevivência**. 2011. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2011. doi:10.11606/T.11.2011.tde-12092011-170753. Acesso em: 2018-12-02.