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

12 de abril de 2024

1 Introduction

An important area of Statistics is Survival Analysis which has as goal analyze times until the event of a specific event event , sometimes with censored observations . The data set is then formed per samples that include time, censored observations and other features that influence the study in question .

The work is composed for the development of a regression model based at Kumaraswamy distribution , using the Weibull function to compose the Kumaraswamy-Weibull distribution due to the fact that Weibull is a distribution with many applications in this study area . In addition of that , it was the concept of cure fraction was used , which happens when it is assumed that part of the population he was cured during the study , and the covariate insertion concepts in one of the parameters of the probability distribution used .

2 Material e Methods

2.1 Material

The dataset are originating from a cohort with 862 cancer patients hospitalized in the Inca ICU, published by Soares et al. (2006), whose main objective of the study he was to assess factors associated with survival .

The features included in the study are : patient 's sex ; age in years complete ; tumor type , categorized in solid localized , metastatic and hematological ; malnutrition patient , recent weight loss above $> 10\%$ or $BMI < 18$; presence of severe comorbidity ; presence of leukopenia

2.2 Methods

To build a regression model at Survival Analysis need to understand how time affects by features included in the model. The Kaplan-Meier estimator is a technique used to describe the time in relation to others variables . The Kaplan-Meier survival curve plot for the data in question do not tends to zero at the end of the curve, which indicates the use of a cure fraction model .

*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

None of the models parametric usual had adjustment acceptable in this data set. Therefore, the best path was to find a way to adjust by another model , through the Kumaraswamy-Weibull distribution with cure fraction .

The Kumaraswamy distribution he has as cumulative distribution function:

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

And probability density function:

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

where $a > 0$, $b > 0$ and $G(t)$ is a CDF arbitrary which can be any probability distribution , and $g(t)$ is its PDF.

The distribution Weibull was chosen given your frequent use in Survival Analysis . $G(t)$ will be the Weibull CDF , which is defined by:

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

where $c > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter .

Then , the PDF of the K-Weibull distribution will be :

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

And the CDF:

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

Given the relation $S(t) = 1 - F(t)$, the survival function of the K -Weibull distribution is given by :

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

To obtain a regression model for the K-Weibull distribution , we need a re-parameterization of two parameters : $\lambda = exp\{-\mu\}$ and $c = \frac{1}{\sigma}$, where $-\infty < \mu < \infty$, and $\sigma > 0$.

When considering the vector of variables regressors $\mathbf{x} = (x_0, x_1, ...x_p)^T$, we consider $\mu = \mathbf{x}^T\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, ...,\ beta_p)^T$ is the unknown parameter associates to the features. In K-Weibull functions we will have :

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

A cure fraction model he has as survival function :

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

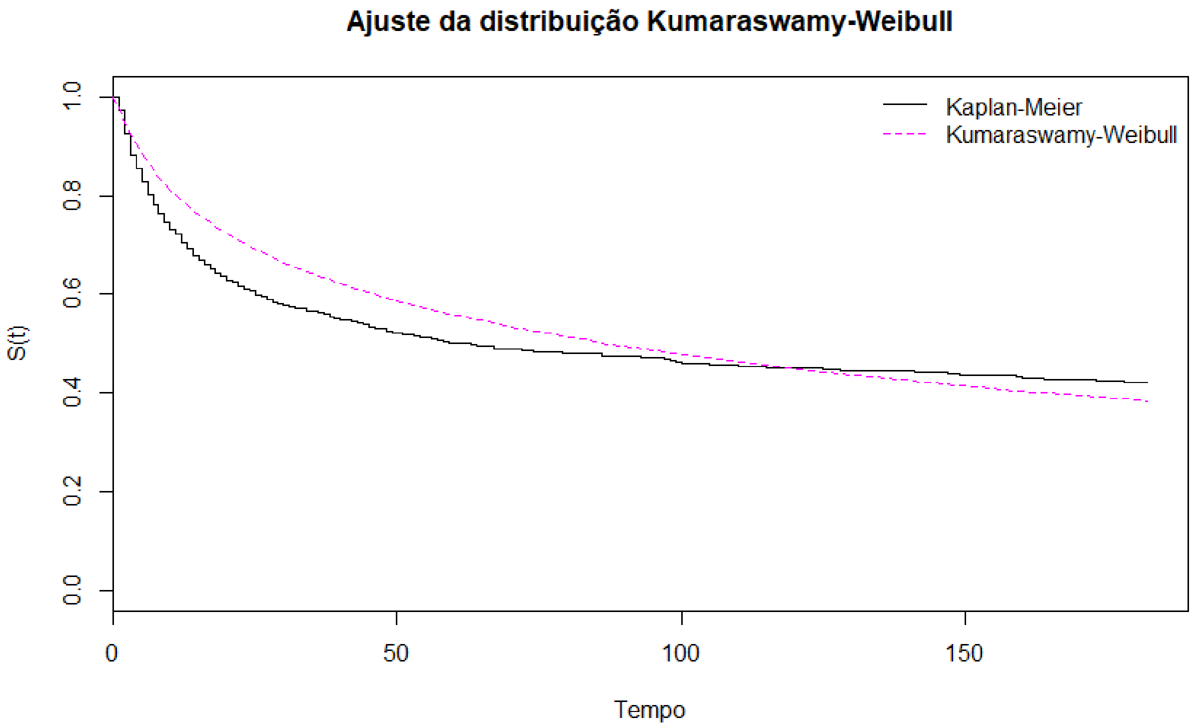
and function probability density :

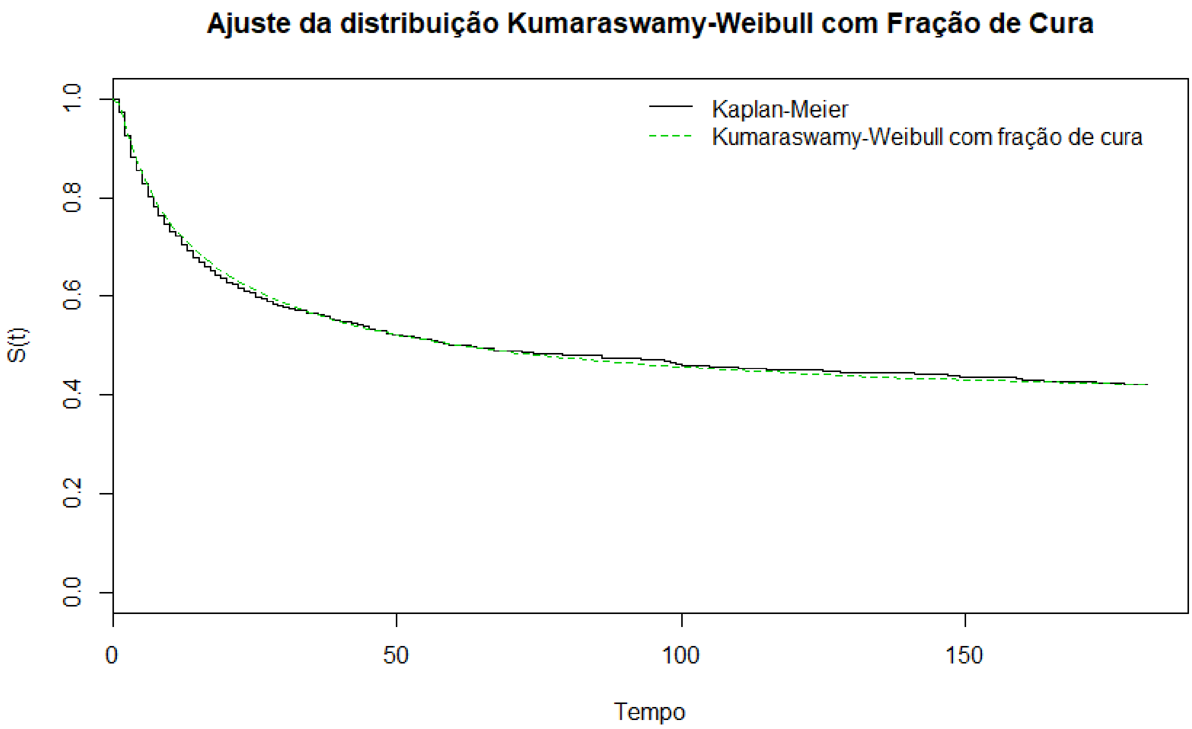
$$f_{pop}(t) = (1 - \phi)f_{mod}(t) \tag{9}$$

where $S_{mod}(t)$ and $f_{mod}(t)$ are the survival function and the pdf of a model respectively. In our case, these are the functions of the K-Weibull model . ϕ is a parameter that represents the fraction of the population that was cured during the study, that means, at the end of the study they were not anymore at risk of failure.

3 Results

The adjustment of the Kumaraswamy-Weibull model, without cure fraction and with cure fraction, respectively:





Both adjustments shown above where made without any features.

To introduce the features, the analysis were made through Kaplan-Meier survival curves for covariate categorical test, Log-Rank and Wilcoxon tests. These helped define the features to be incorporated into the regression model, and through the foward-stepwise approach, the significant variables were included. All features, except age, are categorical ,due to this, it was necessary to input *dummy* variables to introduce the qualitative information correctly.

The final model estimates are the following:

Variable	Estimate	Standard Error	p-value
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	-

Where:

- β_1 refers to the leukopenia variable (X_1)
- β_2 e β_3 are referring to the patient's tumor type (X_2 e X_3)
- β_4 refers to the patient's age variable (X_4)
- β_5 refers to the patient's malnutrition variable (X_5)

The presence of a hematological tumor or Metastatic decreases the patient's probability of survival when compared to the patients who have solid tumor (Located) . Likewise, the presence of leukopenia also decreases the patient's chance of surviving. The older the patient is, the lower is his chances of surviving .

Regarding the presence of signs of malnutrition (recent weight loss above of 10 % or BMI <18) decreases the patient's probability of surviving . Finally , the presence of severe comorbidity has no significant influence at probability of patient surviving , as well as the patient's sex.

4 Conclusion

This work purpose was to develop a regression model and apply a non-trivial distribution function in Survival Analysis with computer methods.

The cure fraction in the regression model based on Kumaraswamy-Weibull distribution provide the adjusted curve to be closer to the survival curve for the data. The results presented were satisfying and they are compatible with the model proposed. It is well explanatory by the chosen function.

5 References

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.