A Distribuição Weibull na Modelagem de Dados de Tempo Até o Divórcio

Fabio Douglas Soares Bezerra[1], Daniel Valentins de Lima[2], Evandro Mariano Barros da Silva[3], Marcelino Alves Rosa de Pascoa[4], Graziela Dutra Rocha Gouvea[5]

Introduçao

A distribuição de Weibull foi proposta originalmente por W. Weibull (1939) em estudos relacionados ao tempo de falha devido a fadiga de metais e, desde então, vem sendo frequentemente usada em estudos biomédicos e industriais. A sua popularidade em aplicações práticas deve-se ao fato dela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função de taxa de falha é monótona, isto é, ela é estritamente crescente, estritamente decrescente ou constante (COLOSIMO; GIOLO, 2006).

Para uma variável aleatória T com distribuição Weibull tem-se a função de densidade de probabilidade, de sobrevivência e de risco dadas, respectivamente por

$$f(t) = \frac{\tau}{\alpha^{\tau}} t^{\tau-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right],$$

$$S(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right]$$

e

$$h(t) = \frac{\tau}{\alpha}\left(\frac{t}{\alpha}\right)^{\tau-1},$$

em que $t > 0$, τ é o parâmetro de forma, e α o de escala, ambos positivos.

Objetivo

O objetivo deste trabalho foi através do método de máxima verossimilhança, estimar os parâmetros da distribuição de Weibull, utilizando dados de tempo de um casamento até o seu divórcio e comparar seu ajuste com o estimador de Kaplan-Meier.

[1] UFMT. email: *email*.fabiodouglas.soares@gmail.com
[2] UFMT. email: *email*.dvalentins@outlook.com
[3] UFMT. email: *email*.evandromarina_barros@live.com
[4] UFMT. email: *email*. marcelino.pascoa@gmail.com
[5] UFOP. email: *email*. gragouvea@gmail.com

Metodologia

Os dados foram obtidos através da verificação dos registros de casamento realizados em um cartório, situado na Cidade de Cuiabá-MT. Foram avaliados 326 casamentos entre os anos de 2014 e 2018. Tendo como objeto de interesse o estudo do tempo até a ocorrência do divórcio.

Uma forma empírica de determinar o comportamento da função risco se da por meio da construção do gráfico do tempo total em teste (curva TTT), proposto por Aarset (1987). A curva TTT é obtida construindo um gráfico de

$$G\left(\frac{r}{n}\right)=\frac{\sum_{i=1}^{r} T_{i:n}+(n-r)T_{r:n}}{\sum_{i=1}^{n} T_{i:n}} \qquad \text{por} \qquad \frac{r}{n},$$

em que n é o tamanho da amostra, $r = 1, \ldots , n$ e $T_{i:n}$, $i = 1, \ldots , n$ são estatísticas de ordem da amostra.

A estimação dos parâmetros foi feita pelo método de máxima verossimilhança. Para que fosse possível realizar inferências fundamentadas no modelo, foi necessário obter a função de verossimilhança, logo considere uma amostra $X = (x_1, \ldots , x_n)$ de n observações independentes, e seja F o conjunto do logaritmo dos tempos de vida e C o conjunto do logaritmo dos tempos de censura. Assim, o logaritmo da função de verossimilhança para o vetor de parâmetros $\theta=(\alpha,\tau)^T$ para o modelo (1) tem a forma $l(\theta)=\sum_{i\in F} l_i(\theta)+\sum_{i\in C} l_i^{(c)}(\theta)$, em que $l_i(\theta)=\log\left[f(x_i)\right]$, $l_i^{(c)}(\theta)=\log\left[S(x_i)\right]$, $f(x_i)$ é a função densidade de probabilidade (1) e $S(x_i)$ é a função de sobrevivência (2). Dessa forma, o logaritmo da função de verossimilhança para θ é

$$l(\theta)=r\log(\tau)-r\tau\log(\alpha)+(\tau-1)\sum_{i\in F} t-\sum_{i\in F}\left(\frac{t}{\alpha}\right)^{\tau}-\sum_{i\in C}\left(\frac{t}{\alpha}\right)^{\tau}$$

Os componentes do vetor score $U=\left(\frac{\partial l}{\partial \alpha},\frac{\partial l}{\partial \tau}\right)$ são obtidos por diferenciação de θ em relação aos parâmetros, logo

$$U_{\alpha}(\theta)=-\frac{r\tau}{\alpha}+\frac{\tau}{\alpha^{\tau+1}}\left(\sum_{i\in F} t^{\tau}+\sum_{i\in C} t^{\tau}\right)$$

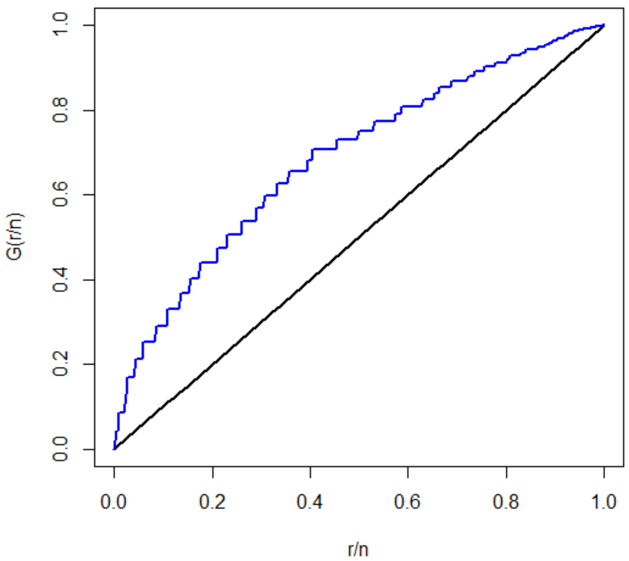
$$U_{\tau}(\theta)=\frac{r}{\tau}-\sum_{i\in F} t-\sum_{i\in F}\log\left(\frac{t}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\tau}-\sum_{i\in C}\log\left(\frac{t}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\tau}$$

As estimativas dos parâmetros são soluções simultâneas das equações $U_{\alpha}(\theta)=0$ e $U_{\tau}(\theta)=0$. As análises foram implementadas no *software* estatístico R.

Resultados e discussões

A Curva TTT para o conjunto de dados de tempo até o divórcio encontra-se na Figura 1 e indica uma função risco na forma crescente. Assim, como a distribuição Weibull modela a função de risco na forma crescente ela é apropriada para analisar esse conjunto de dados.

Figura 1: Curva TTT para dados de divórcio.



Na Tabela 1, podem ser vistos as estimativas de máxima verossimilhança (e os correspondentes erros-padrão que estão entre parênteses) dos parâmetros e os valores das estatísticas dos modelos, Exponencial e Weibull. Os resultados indicam que o modelo Weibull tem os menores valores das estatísticas AIC (Critério de Informação de Akaike), BIC (Critério de Informação Bayesiano) e CAIC (Critério de Informação Akaike Consistente) entre os modelos ajustados, portanto, o modelo Weibull é o mais adequado para os dados de tempo de até o divórcio. O teste da razão de verossimilhança é apresentado na Tabela 2. Os resultados nessa tabela sugerem que o modelo Weibull produz um ajuste mais adequado a esses dados quando comparado com a distribuição exponencial.

Tabela 1: Ajuste final dos modelos comparados, para os dados de tempo até o divórcio

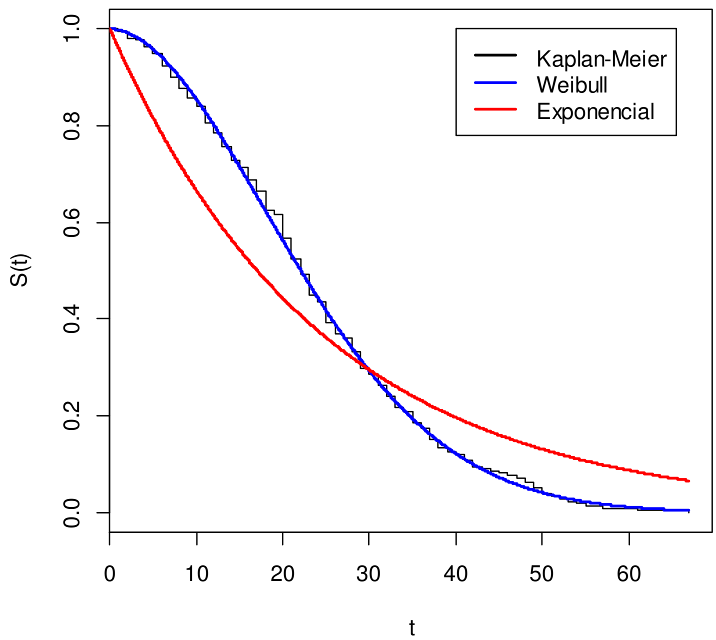
Modelo	α	τ	AIC	CAIC	BIC
Weibull	26,8656	1,8720	2405,3	2405,4	2412,8
	(0,8564)	(0,0842)			
Exponencial	24,5651	1	2547,9	2548,0	2551,7
	(1,4113)	(-)			

Tabela 2: Teste da razão de verossimilhança, para os dados de tempo até o divórcio

Modelo	Estatística do Teste	Valor *p*
Weibull *vs* Exponencial	144,6	< 0,0001

A Figura 2 apresenta a comparação das estimativas da função de sobrevivência segundo Kaplan-Meier e segundo os modelos Weibull e exponencial, para os dados de tempo até o divórcio. Observa-se pela figura que a distribuição Weibull nos fornece um ajuste satisfatório para os dados em estudo. Logo, conclui-se que os casais que não estão mais juntos se divorciam em média 24 meses após o casamento. Após 36 meses aproximadamente 83% dos casais que não estão mais juntos haviam se divorciado.

Figura 2: Estimativas da função de sobrevivência segundo Kaplan-Meier e segundo os modelos Weibull e Exponencial, para os dados de tempo até o divórcio



Conclusão

A distribuição Weibull proposta por Weibull (1954), apresentou melhor ajuste para os dados em estudo, segundo o teste da razão de verossimilhança e as estatísticas AIC, BIC e CAIC, quando comparada com a distribuição exponencial, se mostrando mais flexível.

Agradecimentos

Agradeço primeiramente a Deus e a minha esposa pela motivação a apreço ao estudo, bem como ao professor Marcelino e ao Cartório Xavier de Matos na pessoa da funcionária Eliza Santa que nos concedeu a oportunidade de realizar este trabalho.

Referencias Bibliográficas

AARSET, M. V. How to identify bathtub hazard rate. **Transaction son Reliability**. v. 36, p. 106-108, 1987.

COLOSIMO, E. A.; GIOLO, S.R., **Analise de Sobrevivência Aplicada.**Projeto Fisher – ABE. São Paulo: Edgard Blucher Ltda., 1º Edição 89 p., 2006.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0, URL http://www.R-project.org/.

WEIBULL, W. A. A statistical theory of the strength of material. Royal Swedish Institute for Engineering Research, v. 151, p. 1–45, 1939.