# APLICAÇÃO DO MÉTODO DE BUCKLEY-JAMES COMO ALTERNATIVA AO MODELO DE COX NA VIOLAÇÃO DA PRESSUPOSIÇÃO DE RISCOS PROPORCIONAIS.

Rogério Antonio de OLIVEIRA[1]
Giovanni Faria SILVA[2]
Liciana Vaz de Arruda SILVEIRA[3]

- RESUMO: O método de Buckley-James não tem pressuposição sobre a distribuição dos erros do modelo de regressão linear e pode ser uma alternativa ao modelo semiparamétrico de Cox, quando a pressuposição de proporcionalidade dos riscos não for satisfeita. Como exemplo de aplicação, foram considerados os dados de tempos de vida de pacientes submetidos a um procedimento cirúrgico, utilizado no tratamento de varizes esofágicas. Nesse estudo, foram anotados os tempos até o óbito durante o acompanhamento de 94 pacientes, os que não foram a óbito foram considerados como censuras. Para avaliar o tempo de sobrevivência dos pacientes foram consideradas as covariáveis: classificação do grau da cirrose (Child-Pugh = 1, 2 ou 3), uso do medicamento β-bloqueador e idade do paciente no início do estudo. Como estas covariáveis não apresentaram riscos proporcionais, não foi possível ajustar o Modelo de Cox. Nesse caso, foi ajustado o modelo de regressão linear de Buckley-James (BJ) para dados com tempos de vida censurados. Os valores de p das covariáveis idade, Child = 1, Child = 2 e β-bloqueador = 1 foram, respectivamente, iguais a 0,051; 0,002; 0,056 e 0,023. O modelo de BJ é uma extensão dos modelos lineares clássicos para dados censurados e possui boas propriedades estatísticas sob certas condições de regularidade.

- PALAVRAS-CHAVE:Análise de sobrevivência; regressão para dados censurados; riscos não proporcionais.

## 1 Introdução

O modelo de riscos proporcionais de Cox (1972) é o modelo estatístico mais utilizado no ajuste de dados censurados. Praticamente, todos os pesquisadores têm acesso fácil ao procedimento para o ajuste do modelo semiparamétrico de Cox, pois quase todos os *softwares* estatísticos possuem a implementação do algoritmo.

[1] Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, Instituto de Biociências, Departamento de Bioestatística, Caixa Postal 510, CEP: 18618-970, Botucatu, SP, Brasil. E-mail: *rogerio@ibb.unesp.br*

[2] Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, Faculdade de Medicina de Botucatu, Departamento de Clínica Médica, Caixa Postal 584, CEP: 18618-970, Botucatu, SP, Brasil. E-mail: *giovanni@fmb.unesp.br*

[3] Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, Instituto de Biociências, Departamento de Bioestatística, Caixa Postal 510, CEP: 18618-970, Botucatu, SP, Brasil. E-mail: *liciana@ibb.unesp.br*

A outra razão para a popularidade do modelo de Cox é a facilidade na interpretação dos parâmetros do modelo, pois vários pesquisadores estão já familiarizados com estas análises estatísticas. Entretanto, quando a pressuposição de riscos proporcionais não é satisfeita, o modelo de Cox não pode ser utilizado na modelagem e torna-se necessário utilizar algum outro modelo paramétrico na modelagem dos tempos, por exemplo, o modelo Gama, Log-logística e Log-normal. No entanto, existe também o método de Buckley-James (1979), que é uma análise de regressão de mínimos quadrados adaptada para dados com censuras.

O método de regressão linear de Buckley-James (BJ) possui boas propriedades, pois é consistente sob as usuais condições de regularidade (Lai e Ying, 1991). Diferente do modelo de Cox, o modelo de BJ enfoca na estimação do valor esperado do tempo de sobrevivência, utilizando as variáveis explicativas do modelo.

A grande explicação para a baixa utilização da metodologia de BJ é devido à falta de *softwares* apropriados, pois poucos disponibilizam o procedimento para a estimação dos parâmetros do modelo de regressão linear múltipla para dados com censuras.

## 2 Materiais e métodos

### 2.1 Descrição dos dados

Para ilustrar a aplicação do método de Buckley-James (BJ), foi utilizado o conjunto de dados de sobrevivência de pacientes submetidos a um procedimento cirúrgico para corrigir o problema de varizes no esôfago.

As varizes esofágicas são canais vasculares que unem a circulação venosa porta e sistêmica, ocasionadas por um processo de cirrose hepática. Elas se formam, preferencialmente, na submucosa do esôfago inferior, como consequência do aumento da pressão da veia porta (hipertensão portal), devido a uma complicação progressiva da cirrose hepática.

A cirrose caracteriza-se pela formação de tecido fibrótico que altera a arquitetura hepática levando ao aumento da pressão no sistema porta e, consequentemente, a formação dos trajetos varicosos.

A ligadura elástica de varizes esofágicas (LEVE) é considerada o melhor procedimento para o tratamento das varizes de esôfago, sendo realizado para a prevenção da ruptura desses vasos. O processo de ruptura e sangramento das varizes esofágicas são complicações maiores da hipertensão portal e são acompanhados por uma alta taxa de mortalidade.

Os dados são referentes a 94 pacientes submetidos à LEVE do Hospital de Clínicas da Faculdade de Medicina da UNESP – Campus Botucatu, SP, que foram coletados a partir do ano de 2006. Nesse estudo foram anotados os tempos após o procedimento cirúrgico até o óbito dos pacientes, sendo considerados como tempos censurados os tempos de seguimento em que o paciente não apresentou óbito. Para avaliar o tempo de sobrevivência dos pacientes de LEVE em meses, foram consideradas as seguintes covariáveis: classificação do grau da cirrose (Child-Pugh = 1-estado inicial, 2-intermediário ou 3-avançado), uso do medicamento $\beta$-bloqueador ($\beta$-bloqueador =1 ou 0) e idade do paciente em anos no início do estudo.

## 2.2 Modelo de Cox

Em análise de dados censurados é comum o uso de modelos de regressão. Existem duas classes de modelos de regressão para tempo de vida contínuo: os modelos paramétricos, também chamados de modelos de teste de vida acelerado, e o semiparamétrico, chamado de modelo de riscos proporcionais ou modelo de regressão de Cox (Kalbfleisch & Prentice, 1980).

Os modelos paramétricos podem ser encontrados com detalhes em vários textos, tais como Kalbfleisch & Prentice (1980), Lawless (1982) e Cox & Oakes (1984). Cox (1972) introduziu uma nova dimensão de flexibilidade com a análise de dados censurados com covariáveis. A função de risco no tempo $t$, dado o vetor $\boldsymbol{x}$ = ($x_1$, ...,$x_p$)' de covariáveis, é dada por

$$h(t,\boldsymbol{x}) = h_0(t)\exp\left(\boldsymbol{\beta}'\boldsymbol{x}\right)$$

em que:

$\boldsymbol{\beta}$ é o vetor de parâmetros de regressão a ser estimado;

$h_0(t)$ é a componente não paramétrica, uma função não negativa do tempo que não é especificada, usualmente chamada de função base, pois $h_0(t;\boldsymbol{x}) = h_0(t)$ quando $\boldsymbol{x} = \mathbf{0}$;

$T$ representa a variável resposta, tempo até a ocorrência de um evento.

A fim de estimar $\boldsymbol{\beta}$ para dados censurados, Cox (1975) propôs a função de verossimilhança parcial, que não depende de $h_0(t)$, somente depende de $\boldsymbol{\beta}$**.** Ela é construída para uma amostra de $n$ observações com base na probabilidade condicional da $i$-ésima observação vir a falhar no tempo $t_i$, $i$=1, 2, ..., $n$ e que ocorre uma falha neste tempo. A função de verossimilhança parcial a ser usada para as inferências no modelo de Cox é dada por

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp\left(\boldsymbol{\beta}'\boldsymbol{x_i}\right)}{\sum_{l \epsilon R_i} \exp\left(\boldsymbol{\beta}'\boldsymbol{x_l}\right)} = \prod_{i=1}^{n} \left[\frac{\exp\left(\boldsymbol{\beta}'\boldsymbol{x_i}\right)}{\sum_{l \epsilon R_i} \exp\left(\boldsymbol{\beta}'\boldsymbol{x_l}\right)}\right]^{\delta_i},$$

em que:

r representa o número de falhas distintas na amostra;

$R_i$ denota o conjunto de índices dos indivíduos sob o risco no tempo $t_i$ e

$\delta_i$ é a variável indicadora de falha, isto é, se $\delta_i$=1 para falha e $\delta_i$=0 para censura.

As propriedades assintóticas dos estimadores de máxima verossimilhança parcial são necessárias para construir intervalos de confiança e testar hipóteses sobre os coeficientes do modelo. Vários autores estudaram estas propriedades (Cox, 1975;Oakes, 1977), mas foram Andersen & Gill (1982) que apresentaram as provas mais gerais das propriedades dos estimadores. Eles usaram a relação entre os tempos de falhas com processo de contagem para mostrar que estes estimadores são consistentes e assintoticamente normais sob certas condições de regularidade.

Para testar a proporcionalidade de riscos em modelos de Cox, considere a realização de um experimento controlado que visa comparar o tempo de vida de dois grupos de pacientes: grupo controle (tratamento = 0) e o grupo tratado (tratamento = 1), que possuem funções de taxa de falha iguais $h_0(t)$ e $h_1(t)$, respectivamente. Suponha que o vetor $\boldsymbol{\beta}$ possui apenas uma covariável representada pelo variável tratamento. Quando a suposição de proporcionalidade de riscos é verdadeira, tem-se que:

$$\frac{h_0(t,x)}{h_1(t,x)} = exp(\beta_1),$$

isto é, a razão das taxa de falhas é constante em todo o tempo de seguimento do estudo. Esta suposição é básica para o uso do modelo de regressão de Cox. Entretanto, existem casos em que a suposição de proporcionalidade de riscos não é válida, ocasionando em sérios problemas na estimação dos parâmetros do modelo. A verificação da proporcionalidade dos riscos pode ser verificada utilizando o método da covariável dependente do tempo.

Suponha o modelo de Cox para o exemplo ilustrativo apresentado anteriormente com a inclusão da covariável dependente do tempo, isto é,

$$h(t,x) = h_0(t)\, exp(\beta_1 * tratamento + \beta_2 * tratamento * t).$$

O teste consiste em verificar a significância do parâmetro $\beta_2$ e se sua estimativa for significativa então a razão das taxas de falhas não é constante no tempo, pois a razão depende do tempo, ou seja,

$$\frac{h_0(t,x)}{h_1(t,x)} = exp(\beta_1 + \beta_2 * t).$$

Observe que se $\beta_2 > 0$ então a taxa de falha do tratamento aumenta ao longo do tempo e se $\beta_2 < 0$, então a razão de taxas de falha decresce com o tempo, ou seja, a taxa de falha do tratamento diminui ao longo do tempo de seguimento do estudo.

## 2.3 Método de Buckley-James

O modelo de BJ considera que a variável resposta tempo $T$ está linearmente relacionada com o vetor de covariáveis $\boldsymbol{x}$, então se $T_i$ é tempo de vida do $i$-ésimo paciente e $\boldsymbol{x}_i$ é o vetor de covariáveis do $i$-ésimo paciente, tem-se:

$$T_i = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i + e_i, \quad i = 1, 2, \dots, n$$

em que $e_i$ são erros aleatórios independentes e identicamente distribuídos com $E(e_i) = 0$ e $V(e_i) = \sigma^2 > 0$ e são independentes de $\boldsymbol{x}$. Considerando a censura, observa-se $Y_i = min(T_i, C_i)$, em que $C_i$ são os tempos censurados, desta forma, a abordagem usual de mínimos quadrados não pode ser utilizada diretamente. Entretanto, Buckley-James (1979) apresentaram uma adaptação para utilizar as informações censuradas, como apresentada abaixo. Considere

$$Y_i^* = Y_i\delta_i + E(T_i|T_i > Y_i)(1 - \delta_i),$$

em que $\delta_i = I(T_i \leq C_i)$ é a variável indicadora de falha. Note que $E(Y_i^*) = E(T_i)$. A ideia básica é substituir $Y_i$ para as observações censuradas por $Y_i^*$. Portanto, tem-se que $Y_i^* = T_i$ se $\delta_i = 1$ e, para as observações censuradas, $Y_i^* = E(T_i|T_i > Y_i)$ se $\delta_i = 0$. No entanto, é necessário calcular $E(T_i|T_i > Y_i)$ para cada $Y_i$.

$$E(T_i|T_i > Y_i) = E(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i + e_i|\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i + e_i > Y_i) =$$
$$= \beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i + E(e_i|\, e_i > Y_i - \beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i)) =$$
$$= \beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i + \int_{Y_i-(\beta_0+\boldsymbol{\beta}'\boldsymbol{x}_i)}^{\infty} e\, \frac{dF}{1 - F(Y_i - (\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i))}$$

em que *F* representa a função distribuição acumulada do erro *e*. Substituindo *F* por suas estimativas encontradas pelo estimador de Kaplan-Meyer, tem-se que

$$y_i^* = y_i\delta_i + \left(\boldsymbol{\beta}'\boldsymbol{x_i} + \frac{\sum_{e_j>e_i} w_j e_j}{1-\hat{F}(e_i)}\right)(1 - \delta_i),$$

$w_j$ são os degraus da função estimada $\hat{F}$. Pode-se notar que o método de estimação não faz nenhuma pressuposição da distribuição dos erros do modelo estatístico.

Desta forma, os parâmetros do modelo podem ser estimados como

$$\widehat{\boldsymbol{\beta}} = \frac{(\boldsymbol{X} - \overline{\boldsymbol{x}})'\boldsymbol{y}^*(\widehat{\boldsymbol{\beta}})}{(\boldsymbol{X} - \overline{\boldsymbol{x}})'(\boldsymbol{X} - \overline{\boldsymbol{x}})},$$

em que $\boldsymbol{X}$ é a matriz das covariáveis e $\overline{\boldsymbol{x}}$ é o vetor das médias das covariáveis. Substituindo $y_i^*$ por suas estimativas e considerando que suas estimativas dependem de $\boldsymbol{\beta}$, tem-se um processo iterativo. No final, calcula-se o intercepto do modelo estatístico:

$$\hat{\beta}_0 = \bar{y}^* - \beta'\bar{x}.$$

Após estimar os parâmetros do modelo, estima-se a matriz de variâncias e covariâncias usando o método como proposto originalmente por Buckley e James (1979). Nota-se que os resíduos individuais não podem ser calculados por $T_i - (\hat{\beta}_0 + \widehat{\boldsymbol{\beta}'}\boldsymbol{x}_i)$ porque, geralmente, não estão disponíveis devido à presença de censuras.

## 3 Resultados e discussões

O tempo de seguimento máximo dos pacientes foi de 47 meses, aproximadamente quatro anos. As idades dos 94 pacientes submetidos ao procedimento cirúrgico LEVE variam de 13 a 80 anos. Pouco mais de 70% dos pacientes usavam o medicamento $\beta$-bloqueador (β-bloqueador=1), aproximadamente 54% possui cirrose em estado inicial (Child-Pugh=1), cerca de 23% dos pacientes foram classificados nos estados intermediário (Child-Pugh=2) ou avançado (Child-Pugh=3).

O modelo de regressão de Cox foi ajustado aos dados para verificar se a pressuposição de proporcionalidade dos riscos é válida para o conjunto de dados analisados. Optou-se pelo método da covariável dependente no tempo. Após a estimação dos parâmetros do modelo com a covariável $\beta$-bloqueador (uso=1 e não uso=1), verificou-se que não existe proporcionalidade de riscos porque o parâmetro com a covariável dependente do tempo apresentou $\beta_2$=-0,0922 com valor de $p$ igual a 0,001. Logo o modelo de Cox não é adequado para os dados analisados, pois a razão das taxas de falhas diminui com o tempo, ou seja, a taxa de óbito dos pacientes que usam o medicamento β-bloqueador diminui ao longo do tempo de seguimento do estudo.

O modelo de regressão utilizando o método de BJ pode ser uma boa alternativa para a modelagem dos dados de sobrevivência dos pacientes. Para o ajuste dos dados, considerou-se como referência o Child-Pugh=3. A função bj do *software* R foi utilizada para estimar os parâmetros do modelo de regressão linear multivariada. A Tabela 1 apresenta as estimativas dos parâmetros do modelo de BJ. O tempo de sobrevivência dos pacientes depende da classificação do grau da cirrose (Child–Pugh) e o uso de $\beta$-bloqueador ($\beta$-bloqueador=1). Devido à importância da idade na recuperação dos pacientes, optou-se em manter a covariável idade no modelo, embora o valor de $p$ seja um pouco superior ao nível de significância de 5%.

Tabela 1 - Estimativas do Modelo de Regressão Linear de Buckley-James

| Parâmetro | Estimativa | Desvio padrão | Estatística Wald | Valor de $p$ |
|---|---|---|---|---|
| Intercepto | 5,042 | 0,292 | - | - |
| Idade | -0,067 | 0,005 | 3,82 | 0,051 |
| ChildPugh1 | 2,961 | 0,740 | 9,88 | 0,002 |
| ChildPugh2 | 1,741 | 1,062 | 3,67 | 0,056 |
| β-Bloqueador | 1,665 | 0,392 | 5,16 | 0,023 |

Para verificar o ajuste do modelo, são apresentadas nas Figuras 1 e 2, respectivamente, os gráficos de dispersão dos resíduos padronizados versus idade e resíduos padronizados versus identificação dos voluntários. Pode-se ver pelas Figuras 1 e 2 que não existe nenhuma tendência nos resíduos, pois estão distribuídos aleatoriamente em torno de 0.

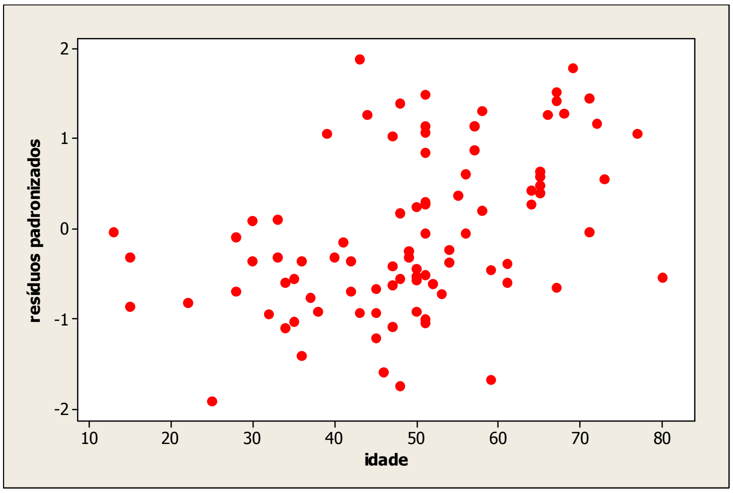


Figura 1 - Gráfico de Dispersão dos Resíduos Padronizados versus idade.

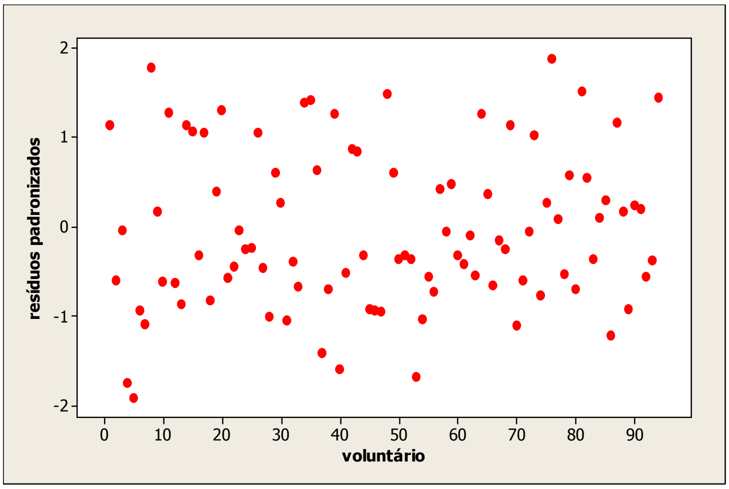


Figura 2 - Gráfico de Dispersão dos Resíduos Padronizados versus identificação voluntários.

## Conclusões

O método de Buckley-James pode ser utilizado para ajustar modelo de regressão linear multivariado de tempo de sobrevivência e, também, é uma alternativa ao modelo de Cox quando não existe proporcionalidade nos riscos. Entretanto, a grande maioria dos *softwares* estatísticos ainda não disponibilizam procedimentos para a estimação dos parâmetros do modelo de regressão pelo método de BJ. Diferente dos modelos paramétricos para dados censurados, Buckley-James (1979) propuseram um estimador que não faz suposição da distribuição dos resíduos. Lai e Ying (1991) mostraram que o modelo de BJ é uma extensão dos modelos lineares clássicos para dados censurados e pode ser uma alternativa quando não há proporcionalidade dos riscos, pois possui boas propriedades estatísticas sob certas condições de regularidade.

OLIVEIRA, R. A.; SILVA, G. F.; SILVEIRA, L. V. A. *Application of Buckley-James method as an alternative to Cox model in violation of proportional risk assumption. Rev. Bras. Biom.* São Paulo, v.33, n.3, p.395-402, 2015.


- *ABSTRACT: The Buckley-James' method (BJ) has no assumptions about the error distribution of the linear regression model. That model can be an alternative to semi-parametric Cox model, when the proportional hazard assumption is not satisfied. For instance, consider the patients' lifetime after a surgical procedure for treating the esophageal variances from a progressive complication of cirrhosis. The lifetimes were observed to death during follow-up of 94 patients. Therefore, patients have not died until the ending of the follow-up were considered as censorship. The patients covariates as degree cirrhosis (Child-Pugh = 1, 2 or 3), use of β-blocker medication and patient age were studied to evaluate their effects in the survival time model. As those covariates presented no proportional hazards, so the Cox model could not be adjusted. In that case, the Buckley-James linear regression model can be useful to analyze data censored* *lifetimes. The covariate p values of age = Child 1, Child = 2 and 1 = β-blocker were respectively equal to 0.051; 0.002; 0.056 and 0.023. The BJ model is an extension of the classic linear models for censored data and it has good statistical properties under certain conditions of regularity.*

▪ *KEYWORDS: Survival Analysis; Regression Model for Censored Data; No Proportional hazards.*

## Referências

ANDERSEN, P.K. & GILL, R. Cox's Regression Model for Counting Processes: A Large Sample Study.*Annals of Statistics*, Pennsylvania, USA, v. 10, p.1100-1200. 1982.

BUCKLEY, J. e JAMES, I. Linear regression with censored data. *Biometrika*, Oxford, England,v.66, p.429-436, 1979.

COX, R. D. Regression models and life-tables (with discussion).*Journal of the Royal Statistical Society*, Series B, London, England, v.34, p.187-220, 1972.

COX, D. R. Partial Likelihood. *Biometrika,* Oxford, England, v.62, p.269-276, 1975.

COX, D. R. & D. OAKES. *Analysis of Survival Data*. 1.ed. London: Chapman and Hall, 1984. 195p.

KALBFLEISCH, J. D. & R. L. PRENTICE. *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons, 1980. 321p.

LAI, T. L., YING, Z., Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *The Annals of Statistics*, Pennsylvania, USA, v.19, n.3, p.1370-1402, 1991.

LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons, 1982. 579p.

OAKES, D. The Asymptotic Information in Censored Survival Data. *Biometrics*, Washington, USA, v.64, p.441-448, 1977.

Recebido em 27.11.2014

Aprovado após revisão em 24.06.2015