Available online at www.alphanumericjournal.com

alphanumeric journal
The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 7, Issue 2, 2019

Received: July 21, 2017
Accepted: December 19, 2018
Published Online: December 31, 2019

AJ ID: 2018.07.02.STAT.05
DOI: 10.17093/alphanumeric.330039
Research Article

# Buckley-James Model in Survival Analysis

Burcu Özcan *

Guelph University, Department of Economics and Finance, Guelph, Ontario, Canada, bozcan@uoguelph.ca

Duru Karasoy, Ph.D.

Prof., Department of Statistics, Hacettepe University, Ankara, Turkey, durdu@hacettepe.edu.tr

* Hacettepe Üniversitesi İstatistik Bölümü, 06800 Beytepe, Ankara, Türkiye

**ABSTRACT**

Cox proportional hazards model is the most common one in survival analysis. Buckley-James model can be offered as an alternative to Cox proportional hazards model. While Buckley-James model is focused on calculation of the expected value of the survival time, Cox proportional hazards model focuses on the relative risk of explanatory variables on the failure event. In this study, Buckley-James model is examined, and is applied on a breast cancer data set. Using this model risk factors affecting survival time of breast cancer is determined.

**Keywords:** Buckley-James Method, Survival Analysis, Cox Proportional Hazards Model, Censored Data

# Yaşam Çözümlemesinde Buckley-James Modeli

**ÖZ**

Yaşam çözümlemesinde en yaygın kullanılan model Cox orantılı tehlikeler modelidir. Cox orantılı tehlikeler modeline alternatif olarak Buckley-James modeli kullanılabilir. Buckley-James modeli yaşam süresinin beklenen değerinin hesaplanmasına odaklı iken Cox orantılı tehlikeler modeli başarısız olaylar üzerinde açıklayıcı değişkenlerin göreli etkilerine odaklıdır. Bu çalışmada, Buckley-James modeli incelenmiş ve meme kanseri verilerine uygulanmıştır. Bu model kullanılarak meme kanserinde yaşam süresini etkileyen risk faktörleri belirlenmiştir.

**Anahtar Kelimeler:** Buckley-James Yöntemi, Yaşam Çözümlemesi, Cox Orantılı Tehlikeler Modeli, Durdurulmuş Veri

# 1. Giriş

Durdurulmuş veriler ile kullanılabilen birçok yöntem geliştirilmiştir. Bu yöntemlerden en çok bilineni ve yaygın olarak kullanılanı Sir David Cox tarafından önerilen Cox orantılı tehlikeler modelidir (Cox, 1972). Bu model, Cox model ya da Cox regresyon modeli olarak da ifade edilmektedir.

Cox orantılı tehlikeler modelinin, yaşam çözümlemesinde yaygın olarak kullanılmasının nedenlerinden ikisi, istatistik paket programlarında kolaylıkla uygulanabilir olması ve bu modelin teorik açıdan avantajları olarak sıralanabilir.

Cox orantılı tehlikeler modeli Eşitlik 1'de verildiği gibidir. Yaşam süresi t'nin özel bir dağılımı olmaması ve temel tehlike fonksiyonu h0(t)'nin özel bir biçimi belirtilmemesine rağmen, 'ların tahmin edilebilir olması da modelin teorik avantajları arasında yer alır.

$$h(\mathrm{t}, \mathbf{x}) = \mathrm{h}_0(\mathrm{t}) \exp\left(\Sigma_{\mathrm{i}}^{k} \beta_{\mathrm{i}} \mathrm{x}_{\mathrm{i}}\right) \qquad (1)$$

Bu model parametrik olmayan bir modeldir. Ancak parametrik ve yarı parametrik doğrusal modellerin kullanılabileceği durumlar ortaya çıkabilir. Bu modellerden bazıları Miller (1976), Buckley ve James (1979) ve Koul vd. (1981) tarafından ele alınmıştır (Glasson, 2007). Durdurulmuş gözlemlerin yaşam süreleri bilinmediği için durdurulmuş verilerle en küçük kareler (EKK) tahmin edicisi hesaplanamaz. Bu üç çalışmada, durdurulmuş veriler kullanılarak EKK prensibine dayanan tahmin yöntemleri geliştirilmiştir. Koul vd. (1981) tarafından önerilen tahmin edici iteratif yöntem gerektirmezken, Miller (1976) ve Buckley ve James'in (1979) önerdikleri iteratif modellerdir (Glasson, 2007). Buckley ve James (1979) tarafından önerilen Buckley-James modeli durdurulmuş verilere uyarlanmış doğrusal en küçük kareler regresyonunu dikkate almaktadır (Buckley ve James, 1979). Stare, Heinzl ve Harrell (2000), Buckley-James modelinin, Cox orantılı tehlikeler modeline tercih edilmesinin nedenlerini; orantılılık varsayımı gerektirmemesi ve istatistikçi olmayanlar tarafından bile kolaylıkla yorumlanabilir olması olarak belirtmişlerdir (Stare vd., 2000). Cui (2005), Buckley-James modelinin yaygın olarak kullanılmamasının nedenini ise uygun bilgisayar programının olmayışı olarak belirtmiştir (Cui, 2005).

Buckley-James yöntemi, Miller (1976) yöntemindeki tutarsızlık sorununun üstesinden gelmektedir (Buckley ve James, 1979). Buckley-James modelini diğer doğrusal regresyon modelleri ile karşılaştıran çalışmalardan biri Miller ve Halpern (1982) çalışmasıdır. Bu çalışmada, Cox (1972), Miller (1976), Buckley ve James (1979) ve Koul vd. (1981) tarafından önerilen dört yöntem ele alınmıştır. Miller ve Halpern (1982), Cox ve Buckley-James modellerinin belli bir durdurma yapısına bağlı olmadığını göstermişler ve güvenilir tahmin ediciler olduğunu kanıtlamışlardır (Miller ve Halpern, 1982). James ve Smith (1984), Buckley-James tahmin edicisinin tutarlılığını incelemişlerdir (James ve Smith, 1984). Kao (1985), durdurulmuş veriler için uygulanan regresyon modellerinde kullanılan Buckley-James ve Schmee-Hahn tahmin edicileri için aykırı değerleri incelemiştir (Kao, 1985). Weissfeld ve Schneider (1987), koşullu bootstrap, Buckley-James ve tanımladıkları tahmin ediciyi bir simülasyon çalışması ile karşılaştırmışlar ve bu üç yöntemin sonuçlarının birbirine benzer olduğunu belirtmişlerdir. Koşullu bootstrap ve önerdikleri tahmin edicinin sonuçlarının Buckley-James'e göre olması gerekenden daha düşük bulduklarını ve katsayıların tahmin edicilerine ait standart sapmalarını ise Buckley-James tahmin edicisine göre daha büyük bulduklarını belirtmişlerdir (Weissfeld ve Schneider, 1987). Moon (1989), Buckley-James modeli ile diğer bir yarı-parametrik doğrusal regresyon modeli olan Tobit yöntemini karşılaştırmıştır. Buckley-James modelinin, Tobit modele göre daha iyi hata kareler ortalamasına (mean squared error) sahip olduğunu belirtmiştir (Moon, 1989). Heller ve Simonoff (1990), Buckley-James yönteminin diğerlerine göre her zaman daha küçük yan ve hata kareler ortalamasına sahip olduğunu ifade etmişlerdir (Heller ve Simonoff, 1990). Lai ve Ying (1991), Kaplan-Meier kestiricisinin hata dağılımının üst kuyruklarında stabil olmaması nedeniyle oluşan zorluklardan kurtulmak için modifiye edilmiş Buckley-James tahmin edicisini kullanmışlardır (Lai ve Ying, 1991). Heller ve Simonoff (1992), Buckley-James modeli ile Cox modeli karşılaştırmışlar ve çoğunlukla, her iki yöntemin benzer performansa sahip olduğu sonucuna ulaşmışlardır. Eğer durdurulmuş veri oranı düşük ise Buckley-James modelini, yüksek ise Cox modelini kullanmayı önermişlerdir. Ortalama seviyede durdurulmuş veri varsa, bu durumda R2 değerine bakmayı tercih etmişlerdir. Durdurulmuş veri kullanım oranı ve Cox model için hesaplanan R2 değerine göre, Cox ile Buckley-James modelleri arasında karar verebilmek için Tablo 1'de verilen tabloyu hazırlamışlardır (Heller ve Simonoff, 1992). Feingold (1993), bu tabloda düzeltme yaparak son satırı eklemiştir (Feingold, 1993).

| **Durdurulmuş veri oranı** | **R2** | **Model seçimi** |
|---|---|---|
| < %40 | Hepsi | Buckley-James |
| %40 - %60 | 0,00 – 0,25<br>0,25 – 0,65<br>0,65 – 1,00 | Buckley-James<br>Cox<br>Buckley-James |
| > % 60 | Hepsi | Cox |
| Düzeltilmiş<br>> %60 | 0,00 – 0,55<br>0,55 – 1,00 | Cox<br>Buckley-James |

**Tablo 1.** Heller-Simonoff 'un ve Feingold'un model seçimi

Lin ve Wei (1992), yaşam verileri ile doğrusal modellerdeki regresyon parametreleri ile ilgili çalışmışlardır (Lin ve Wei, 1992). Hillis (1993), durdurulmuş veriler için uygulanan doğrusal regresyon modelleri arasında, yarı-parametrik varsayım altında regresyon katsayılarını en iyi tahmin eden yöntemin Buckley-James yöntemi olduğunu gösteren bir çalışma yapmıştır (Hillis, 1993). Wu ve Zubovic (1995) doğrusal modeller ile Buckley-James modelini karşılaştıran bir simülasyon çalışması yapmışlardır. Küçük örneklem büyüklüğünde yüksek durdurma oranına sahip olduğunda bile Buckley-James tahmin edicisinin, Tip-II dışındaki durdurma tiplerinde yansızlığını göstermişlerdir (Wu ve Zubovic, 1995). Buckley-James modeli üzerine birçok teorik çalışma yapılmasına rağmen, bu model ile uygulama çalışması gerçekleştirilmemiştir. Bunun başlıca nedeni ise yeterli ve kolay uygulanabilir paket program olmamasıdır. Yu ve Wong (2002) çalışmalarında iteratif olmayan bir algoritma geliştirmişlerdir. Bu algoritmayla bir kaç adımda bütün olası Buckley-James beta tahminleri bulunabilmektedir (Yu ve Wong, 2002). Jin, Lin ve Ying (2006) durdurma üzerindeki zayıf koşullar ve klasik EKK ile karşılaştırılabilir olmasının sonucu olarak, Buckley-James tahmin edicisinin hızlandırmış yaşam modelleri için doğal bir seçim olduğunu belirtmişlerdir. Jin vd. (2006) çalışmalarında, sağdan durdurulmuş veriler ile hızlandırılmış yaşam modelleri için de en küçük kareler prensibini baz alan basit ve güvenilir bir yöntem tanımlamışlardır. Regresyon parametresinin yeni tanımlanan tahmin edicisi, Buckley-James tahmin yöntemine iteratif bir çözümdür. Bu yeni tahmin edicinin tutarlı ve asimptotik olarak normal olduğunu göstermişlerdir (Jin vd., 2006). Wang, Nan, Zhu ve Beer (2008) çalışmalarında, yarı parametrik hızlandırılmış yaşam modelleri için yüksek boyutlu ortak değişkenler ile çift kat cezalandırılmış (doubly penalized) Buckley-James yöntemi geliştirmişlerdir (Wang vd., 2008). Ning, Qin ve Shen (2011), hızlandırılmış yaşam modeli altında, sağdan durdurulmuş yaşam verileri için Buckley-James tipi tahmin edicilerin genelleştirilmesi üzerine çalışmışlardır (Ning vd., 2011). Yu (2011), Buckley-James tahmin edicisini, yarı parametrik ayarlamalar ile genelleştiren bir çalışma yapmıştır (Yu, 2011).

Bu çalışmada Buckley-James modeli incelenmiştir. Daha sonra, 124 hastaya ait meme kanseri verileri kullanılarak Cox orantılı tehlikeler modeli ve Buckley-James modeli ile bir uygulama yapılmış, sonuçlar irdelenmiştir.

## 2. Buckley - James Modeli

Buckley-James modeli, Buckley ve James (1979) tarafından önerilmiş ve durdurulmuş gözlem içeren veriler için kullanılan bir modeldir.

Model, T süresini, x açıklayıcı değişken vektörüyle doğrusal olarak ilişkili varsayar ve Eşitlik 2'deki gibi tanımlanır (Stare vd., 2000):

$$T_i = \alpha + \beta' x_i + \varepsilon_i \qquad (2)$$

$C_i$, (i=1,2,…,n) durdurma süresi olmak üzere $Y_i$ = min($T_i$, $C_i$) olarak ifade edildiğinde Eşitlik 2, $Y_i$ için geçerli olmadığından, EKK yöntemi uygulanamaz (Stare vd., 2000).

Sağdan durdurulmuş veri ile doğrusal regresyon modeli için ($Y_i$, $\delta_i$, $x_i$) ele alınır. $\delta_i$ durum değişkenidir ve

$$\delta_i = \begin{cases} 0, & T_i > C_i \text{ ise} \\ 1, & T_i \leq C_i \text{ ise} \end{cases}$$

biçiminde tanımlanır.

Durdurulmuş veri durumunda EKK yöntemi yanlı tahmin verir. Yanlılığı düzeltmek için literatürde birçok model geliştirilmiştir. Bu modellerden biri de Buckley-James modelidir (Glasson, 2007). Buckley ve James (1979), durdurulmuş veriler yerine onların beklenen değerlerini, $E(T_i|T_i > C_i)$, alarak Eşitlik 3'deki, $Y_i^*$değişkenini tanımlamışlardır:

$$Y_i^* = Y_i\delta_i + E(T_i|T_i > C_i)(1 - \delta_i) \qquad (3)$$

Eşitlik 3'de durum değişkeni $\delta_i = I(T_i \leq C_i)$ olmak üzere, $\delta_i = 1$ ve $\delta_i = 0$ için $E\left(Y_i^*(b)\right) = E(T_i)$'dir. Dolayısıyla, $E(Y_i^*(b)) = E(T_i) = E(\alpha + \beta x_i + \varepsilon_i) = \alpha + \beta x_i$ sonucuna ulaşılır. Durdurulmuş veri kullanıldığı için, Eşitlik 3'deki $E(T_i|T_i > C_i)$değeri bilinmediğinden, $E(T_i|T_i > C_i)$ değeri tahmin edildiğinde,

$$E(T_i|T_i > C_i) = E(\alpha + \beta' x_i + \varepsilon_i|\alpha + \beta' x_i + \varepsilon_i > C_i)$$

$$= \alpha + \beta' x_i + E\left(\varepsilon_i\middle|\varepsilon_i > C_i - (\alpha + \beta' x_i)\right) \qquad (4)$$

elde edilir.

Eşitlik 4'de $E(\varepsilon_i \mid \varepsilon_i > C_i - (\alpha + \beta' x_i))$ ifadesinin tahmin edilmesi için, Eşitlik 4 tekrar düzenlenirse, $\varepsilon_i = T_i - bx_i$ olmak üzere,

$$E(T_i|T_i > Y_i) = \beta' x_i + E\big(\varepsilon_i \big| \varepsilon_i > Y_i - (\beta' x_i)\big)$$

$$E\big(\varepsilon_i \big| \varepsilon_i > C_i - (\beta' x_i)\big) = E(T_i|T_i > C_i) - \beta' x_i$$

$$E\big(\varepsilon_i \big| \varepsilon_i > c_i(b)\big) = E(T_i|T_i > C_i) - \beta' x_i \tag{5}$$

elde edilir.

Eşitlik 5'e göre, $E(T_i|T_i > C_i) - \beta' x_i$'in tahmin edicisi $E(\varepsilon_i|\varepsilon_i > c_i(b))$'dir. Bu nedenle, $\hat{E}_b(\varepsilon_i|\varepsilon_i > c_i(b))$ şeklinde gösterilir.

Yani, $c_i(b) = C_i - bx_i$; $\hat{\beta} = b$ olmak üzere,

$$\hat{E}_b\big(\varepsilon_i \big| \varepsilon_i > c_i(b)\big) = E(T_i|T_i > C_i) - \beta' x_i \tag{6}$$

sonucuna ulaşılır. Elde edilenler, Eşitlik 3'de yerine yazılırsa,

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

$$= Y_i\delta_i + \big[\hat{E}_b\big(\varepsilon_i \big| \varepsilon_i > c_i(b)\big) + \beta' x_i\big](1 - \delta_i)$$

$$Y_i^*(b) = bx_i + \big[\varepsilon_i\delta_i + \widehat{E_b}\big(\varepsilon_i \big| \varepsilon_i > c_i(b)\big)(1 - \delta_i)\big] \tag{7}$$

sonucuna ulaşılır.

Eşitlik 7'deki, $\hat{E}_b\big(\varepsilon_i \big| \varepsilon_i > c_i(b)\big)$ ifadesi aşağıdaki biçimde bulunur:

Gözlemlenmiş artıklar, $e_i(b) = Y_i - bx_i$ ve $e_1(b) < e_2(b) < \cdots < e_n(b)$ olmak üzere,

$$\widehat{E_b}\big(\varepsilon_i \big| \varepsilon_i > c_i(b)\big) = \sum_{k=1}^{n} w_{ik}(b) e_k(b) \tag{8}$$

elde edilir. Eşitlik 8'de, $w_{ik}(b)$ ağırlık fonksiyonudur. $c_i(b)$ fonksiyonundan büyük olan artıkların doğrusal kombinasyonları için ağırlık fonksiyonu, gözlemlenmiş artıklara Kaplan-Meier tahmini (product-limit tahmini) uygulanarak elde edilir:

$$w_{ik}(b) = \begin{cases} \dfrac{d\hat{F}(e_k(b))\delta_k(1-\delta_i)}{\hat{S}(e_i(b))}, & e_k(b) > e_i(b) \text{ ise} \\ 0 & \text{ö.d.} \end{cases} \tag{9}$$

$d\hat{F}_b$, genel Kaplan-Meier tahmini tarafından belirlenen birikimli dağılım fonksiyonudur. $\hat{S}\big(e_i(b)\big)$, $e_i(b)$ artığının Kaplan-Meier tahmin edicisidir. Eşitlik 10'da matris formunda gösterilmiş olan ağırlık matrisinin özellikleri aşağıdaki gibidir (Smith ve Zhang, 1995):

- W idempotent bir matristir ($W^2 = W$ ve $(I - W)^2 = I - W$),
- $1^T W = n d\widehat{F}\big(e_k(b)\big)^T$,
- Ağırlık matrisi, durdurma olmadığında birim matrisi verir.

$$\mathcal{W}(b) = \mathrm{diag}(\delta) + [w_{ik}(b)]$$

ifadesi ağırlık matrisini verir.

Her bir $Y_i^*$ biliniyorsa, $\beta$'nın tahmin edicisi Eşitlik 10 kullanılarak bulunabilir:

$$\sum_{i=1}^{n}(x_i - \bar{x})\big(Y_i^* - \hat{\beta}x_i\big) = 0 \tag{10}$$

$Y_i^*$ yerine, tahmin edicileri yazılarak, iterasyon yöntemi uygulanır. İterasyon sonucunda ise β'nın tahmin edicisi aşağıdaki gibi bulunur:

$$\hat{\beta}^{m+1} = \frac{\Sigma^n (x_i - \bar{x}) Y_i^*(\hat{\beta}^m)}{\Sigma^n (x_i - \bar{x})^2} \quad ; \forall m = 1,2,3,\ldots \tag{11}$$

$\hat{\beta}^m$, $\hat{\beta}$'nın m. iterasyonudur.

$|\hat{\beta}^{m+1} - \hat{\beta}^m| \rightarrow 0$'a yakınsamalıdır. Regresyondaki sabit terim olan α'nın tahmin edicisi ise,

$$\hat{\alpha} = \frac{Y^*(\hat{\beta}) - \hat{\beta} x_i}{n} \tag{12}$$

ile elde edilir.

Buckley-James modeli çoklu regresyon analizinde matris gösterimi ile de yazılabilir. Durdurulmuş veri olduğunda, Eşitlik 13 ile çoklu regresyon modeli için β'nın tahmin edicisi hesaplanabilir:

$$b^{(m+1)} = (X^T X)^{-1} X^T Y^*(b^m) \tag{13}$$

Eşitlik 13'de, $b^m$, b'nin m. iterasyonudur.

$Y = (Y_1, Y_2, \ldots, Y_n)^T$ ve $\delta = (\delta_1, \delta_2, \ldots, \delta_n)^T$ vektörler olmak üzere,

$$Y^*(b) = Xb + \mathcal{W}(b)(Y - Xb) \tag{14}$$

ifadesi Eşitlik 13'de yerine yazılırsa,

$$b^{(m+1)} = (X^T X)^{-1} X^T [Xb^m + \mathcal{W}(b^m)(Y - Xb^m)] \tag{15}$$

elde edilir. Elde edilen, β'nın, Buckley-James tahmin edicisidir ve $\hat{\beta}_{bj}$ ile gösterilir (Glasson, 2007). Stare vd. (2000) çalışmasında, elde edilen Buckley-James tahmin edicisi $\hat{\beta}_{bj}$'nın güvenilir sonuçlar verebilmesi için durdurulmuş veri sayısının düşük olması gerektiğini belirtmişlerdir ((Stare vd., 2000).

## 2.1 Buckley-James Tahmin Edicisinin Varyansı

Buckley ve James (1979), modelin ve katsayıların varyans tahminlerini vermişlerdir. $E_\delta$ ve $var_\delta$ sırasıyla ortalama ve varyans değerleri olmak üzere, Buckley-James modelinin varyansı,

$$var(\hat{\beta}) = E_\delta[var(\hat{\beta}|\delta)] + var_\delta[E(\hat{\beta}|\delta)] \tag{16}$$

biçiminde hesaplanır (Buckley ve James, 1979).

Eşitlik (16)'da, ikinci terimin, birinci terimden daha küçük olduğu varsayıldığında ve Eşitlik (15)'de tanımlanan ağırlık fonksiyonları göz ardı edildiğinde, $var(\hat{\beta})$'nın tahmin edicisi Eşitlik (17)'deki gibi elde edilir (Buckley ve James, 1979):

$$\widehat{var}^*(\hat{\beta}) = \frac{[\Sigma_k var(y_k)\{x_k - \bar{x} + \Sigma^c w_{ik}(\hat{\beta})(x_j - \bar{x})\}^2]}{\Sigma^n (x_i - \bar{x})^2 - \Sigma^c (x_j - \bar{x})\{x_j - \Sigma_k w_{jk}(\hat{\beta}) x_k\}} \tag{17}$$

Eşitlik (17)'de verilen varyans tahmini pratikte doğru sonuç vermediğinden bu formül tekrar düzenlenerek Eşitlik (18) elde edilmiştir (Buckley ve James, 1979):

$$\widehat{var}(\hat{\beta}) = \frac{\{\Sigma^u var(y_k)\,(x_k - \bar{x}_u)^2\}}{\{\Sigma^u (x_k - \bar{x}_u)^2\}^2} \tag{18}$$

Eşitlik (18)'deki varyans tahmini, Eşitlik (17)'deki $\widehat{var}^*(\hat{\beta})$ yardımıyla elde edilir. Ağırlık fonksiyonu $w_{ik}(\hat{\beta})$ yerine, yaklaşığı olan $\frac{\delta_k(1-\delta_j)}{n_u}$ koyularak hesaplanmıştır. $n_u$ durdurulmamış gözlem sayısıdır (Buckley ve James, 1979). Eşitlik (18) matris formatında, $\Delta = \mathrm{diag}(\delta_i)$ olmak üzere,

$$\widehat{var}(\hat{\beta}) = \hat{\sigma}^2(\hat{\beta}) = \hat{\sigma}_{BJ}{}^2(\hat{\beta})(X^T\Delta X)^{-1} \tag{19}$$

ve

$$\hat{\sigma}_{BJ}{}^2(\hat{\beta}) = \frac{\Sigma_i^n \delta_i\left[e_i(\hat{\beta}) - \frac{1}{n_u}\Sigma_j^n\left(\delta_i e_i(\hat{\beta})\right)\right]^2}{n_u - p} \tag{20}$$

şeklinde ifade edilir.

Buckley-James varyans tahmin edicisi üzerine birçok yazar tarafından teorik ve pratik çalışmalar yapılmıştır. Weissfeld ve Schneider (1987), Lin ve Wei (1992) ve Hillis (1993), teorik olarak yeterli bulmasalar da uygulamada birçok durumda Buckley-James varyans tahmin edicisinin iyi ve güvenilir sonuçlar verdiğini göstermişlerdir (Weissfeld ve Schneider, 1987; Lin ve Wei, (1992); Hillis, 1993).

## 3. Uygulama

Bu çalışmada, Ocak 1980 ile Eylül 1991 tarihleri arasında Hacettepe Hastanesi Onkoloji Bölümü'nde meme kanseri tanısı konan 124 hastaya ait veriler kullanılmıştır (Erdoğan, 1993). Çözümleme için STATA 14, IBM SPSS Statistics 23 ve R programları kullanılmıştır. Çalışmada açıklayıcı değişken olarak yaş, menapoz, tümör çapı, aksiller lenf nodu, radyoterapi, toksidasyon, müdahale tipi, hastalığın evresi ve tedavi türü değişkenleri incelenmiştir. Meme kanserinden ölümler olay olarak alınmış, yaşayanlar ise durdurulmuş olarak kabul edilmiştir. 124 hastanın 80'ni (%64,5) ölmüş, 44'ü (%35,5) ise çalışmanın sonunda yaşamını sürdürmüştür. 5 yıllık Kaplan-Meier yaşam olasılığı %59,4'dür. Ortanca yaşam süresi ise 81±6'dır. Uygulamada kullanılan açıklayıcı değişkenlere ait bilgiler Tablo 2'de verilmiştir.

| **Değişken** | ***Değişken Düzeyleri*** | **n** | **%** |
|---|---|---|---|
| Yaş | 50,6±1.1 | 124 | 100 |
| Menapoz | 1. Kesilmiş | 66 | 53,2 |
| | 2. Devam ediyor | 58 | 46,8 |
| Tümör Çapı | 1. < 2 cm | 31 | 25,0 |
| | 2. 2 – 5 cm | 66 | 53,2 |
| | 3. > 5 cm | 27 | 21,8 |
| Aksiller Lenf Nodu | 1. Yok | 50 | 40,3 |
| | 2. 1 – 3 adet | 52 | 41,9 |
| | 3. 4 + | 22 | 17,7 |
| Radyoterapi | 1. Almış | 66 | 53,2 |
| | 2. Almamış | 58 | 46,8 |
| Toksidasyon | 1. Herhangi birşey yok | 76 | 61,3 |
| | 2. Bulantı, kusma, sıcak basması | 48 | 38,7 |
| Müdahale Tipi | 1. Sabit mastektopi | 21 | 16,9 |
| | 2. Modifiye radike mastektopi | 48 | 38,7 |
| | 3. Radike mastektopi | 55 | 44,4 |
| Evre | 1. Evre 1 | 42 | 33,9 |
| | 2. Evre 2 | 82 | 66,1 |
| Tedavi Türü | 1. TMX | 59 | 47,6 |
| | 2. L-PAM | 65 | 52,4 |

**Tablo 2.** Değişkenler ve düzeyleri

Orantılı tehlikeler varsayımının sağlanıp sağlanmadığı Schoenfeld (1982) tarafından önerilen Schoenfeld artıkları kullanılarak test edilmiş ve tüm değişkenler için p>0,05 bulunduğundan değişkenlerin orantılı tehlikeler varsayımını sağladığı görülmüştür (Schoenfeld, 1982). Ancak orantılı tehlikeler varsayımı grafik yöntemle de incelendiğinde (t'ye karşı ln(-lnS(t) grafiği) aksiller lenf nodu, toksidasyon ve müdahale tipi değişkenlerinde düzeylerin çakıştığı, tüm düzeylerin birbirine paralel olmadığı görülmüştür. Bu nedenle orantılı tehlikeler varsayımı sağlanmamaktadır.

Çalışmada öncelikle Cox orantılı tehlikeler modeli uygulanmış ve sonuçları Tablo 3'de verilmiştir. Daha sonra ise Bucley-James modeli uygulanmış ve sonuçları Tablo 4'de verilmiştir.

| Değişken | $\hat{\beta}$ | Standart hata ($\hat{\beta}$) | exp($\hat{\beta}$) | p |
|---|---|---|---|---|
| Yaş | 0,043 | 0,018 | 1,04 | 0,0022 |
| Menapoz | | | | |
| 1. Kesilmiş | Referans | | 1 | |
| 2. Devam ediyor | 0,12 | 0,396 | 1,12 | 0,762 |
| Tümör çapı | | | | |
| 1. <2 cm | Referans | | 1 | |
| 2. 2 – 5 cm | -0,97 | 0,311 | 0,38 | 0,002 |
| 3. >5 cm | -1,12 | 0,397 | 0,33 | 0,005 |
| Aksiller lenf nodu | | | | |
| 1. Yok | Referans | | 1 | |
| 2. 1 – 3 adet | -0,51 | 0,297 | 0,59 | 0,084 |
| 3. 4 + | 0,35 | 0,398 | 1,41 | 0,383 |
| Radyoterapi | | | | |
| 1. Almış | Referans | | 1 | |
| 2. Almamış | -0,37 | 0,245 | 0,69 | 0,127 |
| Toksidasyon | | | | |
| 1. Herhangi birşey yok | Referans | | 1 | |
| 2. Bulantı, kusma, sıcak basması | -0,21 | 0,267 | 0,81 | 0,434 |
| Müdahale tipi | | | | |
| 1. Sabit mastektopi | Referans | | 1 | |
| 2. Modifiye radike mastektopi | 1,02 | 0,419 | 2,79 | 0,015 |
| 3. Radike mastektopi | 0,43 | 0,409 | 1,54 | 0,289 |
| Evre | | | | |
| 1. Evre 1 | Referans | | 1 | |
| 2. Evre 2 | 0,33 | 0,274 | 1,38 | 0,239 |
| Tedavi türü | | | | |
| 1. TMX | Referans | | 1 | |
| 2. L-PAM | 0,09 | 0,291 | 1,09 | 0,764 |

**Tablo 3.** Cox orantılı tehlikeler modeli sonuçları

Tablo 3 incelendiğinde, yaş, tümör çapı ve müdahale tipi değişkenlerinin meme kanserinden ölüm riskini etkileyen faktörler olduğu söylenebilmektedir.

Miller ve Halpern (1982), yaşam süresi değişkeninin 10 tabanına göre logaritmasını alarak kullanmışlardır (Miller ve Halpern, 1982). Heller ve Simonoff'a (1992) göre, yaşam çözümlemesindeki Eşitlik (1)'deki gibi regresyon modelleri çoğunlukla doğrusal değil, çarpımsal olması nedeniyle Buckley-James tahmin edicisi kullanırken de yaşam süresinin logaritması alınarak kullanılmalıdır (Heller ve Simonoff, 1992). Uygulamada benzer yöntem izlenmiştir.

| Değişken | $\hat{\beta}$ | Standart hata ($\hat{\beta}$) | exp($\hat{\beta}$) | p |
|---|---|---|---|---|
| Sabit | 1,86 | 0,62 | | 0,003 |
| Yaş | 0,007 | 0,009 | 1,01 | 0,44 |
| Menapoz | | | | |
| 1. Kesilmiş | Referans | | 1 | |
| 2. Devam ediyor | 0,006 | 0,23 | 1,01 | 0,98 |
| Tümör çapı | | | | |
| 1. <2 cm | Referans | | 1 | |
| 2. 2 – 5 cm | 0,11 | 0,17 | 1,12 | 0,52 |
| 3. >5 cm | -0,17 | 0,21 | 0,84 | 0,42 |
| Aksiller lenf nodu | | | | |
| 1. Yok | Referans | | 1 | |
| 2. 1 – 3 adet | 0,13 | 0,18 | 1,14 | 0,46 |
| 3. 4 + | -0,08 | 0,22 | 0,92 | 0,72 |
| Radyoterapi | | | | |
| 1. Almış | Referans | | 1 | |
| 2. Almamış | -0,30 | 0,13 | 0,74 | 0,02 |
| Toksidasyon | | | | |
| 1. Herhangi birşey yok | Referans | | 1 | |
| 2. Bulantı, kusma, sıcak basması | -0,27 | 0,15 | 0,76 | 0,07 |
| Müdahale tipi | | | | |
| 1. Sabit mastektopi | Referans | | 1 | |
| 2. Modifiye radike mastektopi | 0,07 | 0,20 | 1,07 | 0,75 |
| 3. Radike mastektopi | 0,26 | 0,21 | 1,30 | 0,22 |
| Evre | | | | |
| 1. Evre 1 | Referans | | 1 | |
| 2. Evre 2 | -0,32 | 0,14 | 0,73 | 0,02 |
| Tedavi türü | | | | |
| 1. TMX | Referans | | 1 | |
| 2. L-PAM | 0,13 | 0,17 | 1,14 | 0,45 |

**Tablo 4.** Buckley-James modeli sonuçları

Buckley-James modeline göre; radyoterapi ve evre değişkenleri anlamlı ($p<0,05$) bulunmuştur. Radyoterapi almamış olanların yaşam süresi, almış olanlara göre 1,35 (1/0,74) kat daha kısadır. Evre 2 olanların yaşam süresi ise Evre 1 olanlara göre 1,37 kat daha kısadır.

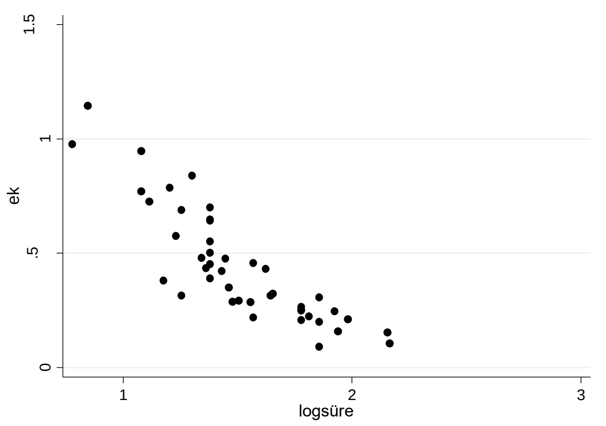
Bu veri setinde durdurulmuş gözlem oranı %35,5'dir. Heller ve Simonoff'un (1992) önerdikleri Cox orantılı tehlikeler ve Buckley-James modelleri arasında karar verme tablosuna göre (Tablo 1) uygun model Buckley-James modelidir. Buckley-James modeli sonucunda elde edilen katsayıların standart hataları, Cox orantılı tehlikeler modeli sonucunda elde edilenlerden daha küçük bulunmuştur. Ayrıca orantılı tehlikeler varsayımı sağlanmadığından bu veriler için Cox orantılı tehlikeler modeli uygun değildir.

Buckley-James modeli ile meme kanseri için beklenen yaşam süresi bulunabilmektedir. Tablo 5, ölçülen süre değişkeni ile tahmini süre değişkenine ait özet istatistikleri vermektedir. Başarısız olanlarda her ikisine ilişkin istatistikler aynıdır, fakat durdurulmuş olanlarda (yaşayanlarda) ortalama tahmini süre daha yüksek fakat standart sapması daha düşüktür.

| Durum | Değişken | n | Ortalama | Standart sapma | Minimum | Maksimum |
|---|---|---|---|---|---|---|
| **Durdurulmuş** | Gözlenen süre | 44 | 1,51 | 0,32 | 0,78 | 2,16 |
| | Tahmini süre | 44 | 1,95 | 0,16 | 1,56 | 2,31 |
| **Başarısız** | Gözlenen süre | 80 | 1,71 | 0,32 | 0,85 | 2,12 |
| | Tahmini süre | 80 | 1,71 | 0,32 | 0,85 | 2,12 |
| **Toplam** | Gözlenen süre | 124 | 1,64 | 0,33 | 0,78 | 2,16 |
| | Tahmini süre | 124 | 1,80 | 0,30 | 0,85 | 2,31 |

**Tablo 5.** Gözlenen ve beklenen süre değişkenine için özet istatistikler

Durdurulmuş gözlemlerde, Buckley-James yöntemi ile gözlenen süreye ne kadar bir ek geldiğini görmek için Şekil 1 elde edilmiştir. Şekil 1 incelendiğinde süre arttıkça gelen ekin küçüldüğü görülmektedir. Gelen ekin ortalaması 0,59, ortancası 0,60, standart sapması 0,25 dir.




**Şekil 1.** Gözlenen süreye karşılık gelen ekler için saçılım grafiği

## 4. Sonuç

Buckley-James modeli, durdurulmuş verilerle kullanılabilen bir modeldir. Yarı-parametrik bir modeldir ve iteratif bir algoritmaya sahiptir. Birçok çalışma sonucunda, Buckley-James modeli, diğer durdurulmuş veriler ile çalışan doğrusal regresyon yöntemleri arasında en iyi performansı gösterdiği belirtilmiştir. Ayrıca öngörü gücü açısından, Cox orantılı tehlikeler modeli ile karşılaştırılabilecek kadar iyi bir model olduğu belirtilmiştir. Buckley-James modeli, yaşam çözümlemesinde beklenen süreyi hesaplamada kullanılabilmektedir. Literatürde çok fazla uygulaması olmamasının nedeni paket programlarda yer almamasıdır. Cox orantılı tehlikeler modeli ise birçok paket programda bulunmakta ve bu nedenle de yaşam çözümlemesinde orantılı tehlikeler varsayımının sağlanıp sağlanmadığına bakılmaksızın çok sık kullanılmaktadır. Ancak, orantılı tehlikeler varsayımının sağlanmadığı durumda Cox orantılı tehlikeler modelini uygulamak ve yorumlamak doğru değildir. Beckley-James modeli, Cox orantılı tehlikeler modeline alternatif bir modeldir ve bu varsayımın sağlanması gerekmemektedir. Ayrıca, varsayımın sağlandığı durumlarda bile durdurma oranının düşük olduğu çalışmalarda bu modelin kullanılması önerilmektedir.

Çalışmada, meme kanseri hastalarına ait verilere Cox orantılı tehlikeler modeli ve Buckley-James modeli uygulanmış ve elde edilen sonuçlar yorumlanmıştır. Bu veriler için durdurma oranı %35,5 olduğundan ve orantılı tehlikeler varsayımı sağlanmadığından Buckley-James modelinin uygun olduğu sonucuna varılmıştır. Buckley-James yöntemi ile meme kanseri hastalarında yaşam süresini etkileyen faktörler olarak radyoterapi ve evre değişkenleri bulunmuştur. Radyoterapi almayanların alanlara göre yaşam sürelerinin 1,35 kat daha kısa, evre 2 olanların evre 1 olanlara göre yaşam sürelerinin 1,37 kat daha kısa olduğu sonucuna varılmıştır.

## Kaynakça

Buckley, J. and James, I. (1979). Linear Regression with Censored Data, Biometrika, 66, 3, 429-436.

Cox, D.R. (1972). Regression Models and Life-Tables, Journal of the Royal Statistical Society, Series B, 34, 2, 187-220.

Cui, J. (2005). Buckley–James Method for Analyzing Censored Data, with an Aplication to a Cardiovascular Disease and an HIV/AIDS Study, The Stata Journal, 5, 4, 517–526.

Erdoğan, A., (1993). Orantılı Hazard Modeli, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Istatistik Anabilim Dalı, Bilim Uzmanlığı Tezi, Ankara.

Feingold, M. (1993). Choice of Prediction Estimator in Censored Regression Models, Biometrics, 49, 661-664.

Glasson, S. (2007). Censored Regression Techniques for Credit Scoring, Phd. Thesis In Mathematics And Computing, Rmit University, 25-44.

Heller, G. and Simonoff, J. S. (1990). A Comparison of Estimators for Regression with a Censored Response Variable, Biometrics, 77, 515-520.

Heller, G. and Simonoff, J. S., (1992). Proportional Hazards and Linear Regression Models, Biometrics, 48, 1, 101-113.

Hillis, S. L. (1993). A Comparision of Three Buckley-James Variance Estimators, Communication in Statistics-Simulation and Computation, 22, 4, 955-973.

James, I. R. and Smith, P. J. (1984). Consistency Results for Linear Regression with Censored Data, The Annals of Statistics, 12, 2, 590-600.

Jin, Z., Lin, D.Y. and Ying, Z. (2006). On Least Squares Regression with Censored Data, Biometrika, 93, 147-161.

Kao, C. (1985). Influence Diagnostics for Censored Regression Models, Statistics and Probability Letters, 3, 337-342.

Lai, T. L. and Ying, Z. (1991). Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis with Censored Data, The Annals of Statistics, 19, 3, 1370-1402.

Lin, J. S. and Wei, L. J. (1992). Linear Regression Analysis Based on Buckley James Estimating Equation, Biometrics, 48, 679-681.

Miller, R. and Halpern, J. (1982). Regression with Censored Data, Biometrika, 69, 3, 521- 531.

Moon, C. (1989). A Monte Carlo Comparison of Semiparametric Tobit Estimators, Journal of Applied Econometrics, 4, 361–382.

Ning, J., Qin, J. and Shen, Y. (2011). Buckley-James Type Estimator with Right Censored and Length–Biased Data, Biometrics, 67, 1369-1378.

Schoenfeld, D. (1982). Partial Residuals for the Proportional Hazards Regression Model, Biometrika, 69, 239-241.

Smith, P. J. and Zhang, J. (1995). Renovated Scatterplots for Censored Data, Biometrika Trust, 82, 2, 447-452.

Stare, J., Heinzl, H. and Harrell, F. (2000). On the Use of Buckley and James Least Squares Regression for Survival Data, New Approaches in Applied Statistics, 16, Metodoloˇ Ski Zvezki, Ljubljana: Fdv, 125-134.

Wang, S., Nan, B., Zhu, J. and Beer, D. G. (2008). Doubly Penalized Buckley-James Method for Survival Data with High-Dimensional Covariates, Biometrics, 64, 132-140.

Weissfeld, L. A. and Schneider, H. (1987). Inferences based on the Buckley-James procedure, Communications in Statistics, Theory and Methods, 16, 6, 1773-87, (1987).

Wu, C. P. and Zubovic, Y. (1995). A Large-Scale Monte Carlo Study of the Buckley-James Estimator with Censored Data, Journal of Statistical Computation and Simulation, 51, 97–119.

Yu, Q. and Wong, G. Y. C. (2002). How to Find all Buckley-James Estimates Instead of Just One?, Journal of Statistical Computation and Simulation, 72, 6, 451-560.

Yu, M. (2011). Buckley-James Type Estimator for Censored Data with Covariates Missing by Design, Scandinavian Journal of Statistics, 38, 252-267.