

## Lojistik ve Cox Regresyon Modellerinin İncelenmesi ve Karşılaştırılması

Durdu SERTKAYA\*

### ÖZET

*Bu çalışmada, epidemiyolojik çalışmalarda herhangi bir olayın risk faktörlerini belirlemek için kullanılan lojistik regresyon ve Cox regresyon modelleri ele alınmıştır. Bu modeller hakkında bilgi verilmiş, aralarındaki matematiksel ilişkiden ve karşılaştırmalardan bahsedilmiştir. Mide kanseri verileri kullanılarak bir uygulama yapılmıştır.*

*Anahtar Kelimeler:* Lojistik Regresyon, Cox Regresyon, Yaşam Çözümlemesi

### 1. GİRİŞ

Son yıllarda epidemiyolojik çalışmalarda veri analizlerinde birçok farklı yaklaşımlar kullanılmaktadır (Green ve Symons, 1983). Son on yıldır klinik çalışmalarda herhangi bir olayın risk faktörlerini belirlemek için göreceli risk (relative risk) regresyon modelleri vazgeçilmez araçlar olmuşlardır. Klinik araştırmalarda hala birbiriyle yarışan iki regresyon modeli vardır. Bunlar lojistik regresyon modeli ve Cox regresyon modelidir. Risk faktörlerini belirlemede bu iki modelden hangisinin daha uygun olduğu hala tartışılmaktadır. Risk faktörlerini tanımlamada, tıp literatüründe görülen istatistiksel analizlerin çoğu lojistik modele dayanmaktadır. Cox regresyon modeli ise onkoloji ve hematoloji gibi tıp alanlarında daha çok görülmektedir (Chevret, 2001). Lojistik regresyon ve Cox regresyon modelleri kronik hastalık uygulamalarında risk faktörlerini belirlemek için epidemiyolojik çalışmaların analizinde kullanılmaktadır. Green ve Symons (1983) bu iki model arasındaki teorik ilişkiyi ortaya koymuşlardır (Green ve Symons, 1983). Gerçekte, lojistik ve Cox regresyonlarının her ikisi de log-doğrusal olarak, belirlenmiş bir olay için risk faktörlerinin kümesiyle ilgili olarak matematiksel modelleme yaklaşımlarıdır (Chevret, 2001). Her iki yöntemde, risk faktörleri kümesi ile belirlenmiş bir olay (genellikle ölüm ya da hastalığın ortaya çıkması) arasındaki ilişkiyi değerlendirmektedir (Ingram ve Kleinman, 1989).

Tıp literatüründe risk faktörlerini belirlemede daha çok lojistik regresyon modeli ile karşılaşılmaktadır. Oysa ki, özellikle epidemiyolojik çalışmalarda hastalar uzun süre takip edildiğinden Cox regresyon modeli daha uygun olmaktadır. Çünkü lojistik

\* Öğr. Gör. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe-Ankara.  
durdu@hacettepe.edu.tr

modelde izleme süresi sabit varsayılmaktadır. Bu durumda da gözlem farklılıkları dikkate alınmamaktadır.

Bu çalışmanın amacı, lojistik ve Cox regresyon modellerini tanıtmak, aralarındaki matematiksel ilişkiyi vermek ve bu iki modeli karşılaştırmaktır.

## 2. REGRESYON MODELLERİ

### 2.1. Lojistik Regresyon Modeli

Lojistik regresyon, bağımlı değişkenin ikili (binary, dichotomous), üçlü ve çoklu kategorilerde gözlemlendiği durumlarda açıklayıcı değişkenlerle neden-sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Açıklayıcı değişkenlere göre yanıt değişkeninin beklenen değerinin olasılık olarak elde edildiği bir regresyon yöntemidir. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilerek risk faktörlerinin olasılık olarak belirlenmesi sağlanmaktadır (Özdamar, 1997).

İkili yanıt değişkeni (bağımlı değişken) Y,

$$Y = \begin{cases} 1, & \text{izleme periyodu boyunca olay meydana gelmişse} \\ 0, & \text{ö.d.} \end{cases} \quad (1)$$

biçiminde tanımlansın. Lojistik model,

$$P(Y | \mathbf{z}, T) = \{1 + \exp[-(\beta_0 + \sum \beta_i z_i)]\}^{-1}$$

biçiminde olmaktadır. Burada  $z_i$  ( $i=1,2,\dots,r$ ) açıklayıcı değişkenler, T, bütün bireyler için aynı olduğu varsayılan toplam izleme uzunluğu,  $\beta_0$  sabit terim ve  $\beta_i$  ( $i=1,2,\dots,r$ ) regresyon katsayıları olarak ifade edilmektedir. Buradan

$$\frac{P(Y = 1 | \mathbf{z}, T)}{P(Y = 0 | \mathbf{z}, T)} = \exp(\beta_0 + \sum \beta_i z_i) \quad (2)$$

olmaktadır. İkili açıklayıcı değişken durumu için (2) eşitliği

$$\frac{P(Y = 1 | \mathbf{z}, T)}{P(Y = 0 | \mathbf{z}, T)} = \begin{cases} e^{\beta_0} & \mathbf{z} = 0 \text{ ise} \\ e^{\beta_0 + \beta} & \mathbf{z} = 1 \text{ ise} \end{cases} \quad (3)$$

ifadesine indirgenmekte ve odds oranı  $e^\beta$  olmaktadır.

(2) ve (3) eşitliklerinden de görüldüğü gibi lojistik model ölüm süresini ve farklı izleme uzunluklarını dikkate almamaktadır. Farklı izleme uzunlukları olduğunda iki yaklaşım uygulanabilmektedir. İlki, eğer farklı izleme uzunluklarının parametre tahminleri üzerine etkilerinin az oldukları varsayılırsa, izleme süreleri dikkate alınmadan bütün bireyler analize dahil edilebilir. Diğer bir yaklaşım ise, T izleme

süresinden daha küçük olan yani yaşayanları analizden çıkarmaktır. Bu yaklaşımların her ikisi de odds oranının yanlış tahmini ile sonuçlanabilmektedir (Ingram ve Kleinman, 1989).

## 2.2. Cox Regresyon Modeli

Birçok durumda incelenen yaşam süresinin başka faktörler tarafından da etkilenebileceği gözönünde bulundurulduğunda, bağımlı değişken olan yaşam süresi üzerinde açıklayıcı değişkenlerin de etkilerinin modellendiği regresyon modelleri yaşam çözümlemesinde önemli bir yer tutmaktadır. Orantılı hazard modeli olarak da ifade edilen bu modeller ilk kez Cox (1972) tarafından ele alınmıştır (Lawless, 1982).

Cox (1972)'un incelediği model

$$h(t, \mathbf{z}) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}) \quad (4)$$

biçimindedir. Burada  $\mathbf{z}$ , açıklayıcı değişkenler vektörü,  $\boldsymbol{\beta}$ , regresyon katsayıları (bilinmeyen katsayılar) vektörü ve  $h_0(t)$ ,  $\mathbf{z} = 0$  olan bir birimin temel hazard fonksiyonu olmaktadır (Cox, 1972; Cox ve Oakes, 1984).  $h(t, \mathbf{z})$ ,  $\mathbf{z}$  açıklayıcı değişkenlerle  $t$  zamanında bir birey için hazard hızını ifade etmektedir.  $h_0(t)$  için özel bir biçim varsayımı yapılmamakta ancak farklı açıklayıcı değişken kümeleri için hazardların orantılı oldukları varsayılmaktadır.  $\boldsymbol{\beta}$ 'nin tahmini

$$L(\boldsymbol{\beta}) = \prod_{k=1}^K \frac{\exp(\boldsymbol{\beta} z_k)}{\sum_{j \in R_i} \exp(\boldsymbol{\beta} z_j)}$$

kısmi olabilirlik fonksiyonu yardımıyla elde edilebilmektedir. Burada  $K$ , ortaya çıkan olay sayısını,  $R_i$  ise  $i$ . olayın meydana gelme zamanında riskte olan bireylerin kümesini göstermektedir (Kalbfleisch ve Prentice, 1980; Ingram ve Kleinman, 1989; Schemper, 1992). Cox regresyon modeli, risk kümesinden ayrılan bireyler tarafından tamamlanmış izleme için de uygun olmaktadır.

İkili açıklayıcı değişken için (4) eşitliği,

$$h(t, \mathbf{z}) = \begin{cases} h_0(t) & \mathbf{z} = 0 \text{ ise} \\ h_0(t) e^{\boldsymbol{\beta}} & \mathbf{z} = 1 \text{ ise} \end{cases}$$

biçimine indirgenmektedir. Hazard oranı ya da görel risk ise  $e^{\boldsymbol{\beta}}$  biçiminde ifade edilmektedir (Ingram ve Kleinman, 1989). Lojistik regresyondaki odds oranı ve Cox regresyondaki görel risk, elde edilen regresyon modelinin veri kümesini ne kadar iyi açıkladığını değerlendirebilmek için gerekli olan özet ölçümlerdir. Görel risk ve odds oranı, önemli bulunan değişken düzeyinin referans kategorisine göre kaç kat daha riskli olduğu yorumunu vermektedir (Hosmer ve Lemeshow, 1999).

### 3. LOJİSTİK ve COX REGRESYON MODELLERİ ARASINDAKİ İLİŞKİ VE BU MODELLERİN KARŞILAŞTIRILMASI

N tane birey olsun. r tane risk faktörü  $\mathbf{z} = (z_1, z_2, \dots, z_r)$ , herbiri T uzunluğundaki izleme periyodunun başlangıcında ölçülmüş olsun. İkili yanıt değişkeni Y, (1) eşitliğindeki gibi tanımlansın. Bu durumda  $\beta_0$  ve  $\beta$  katsayıları ile  $\mathbf{z}$  risk faktörlerinin çoklu lojistik fonksiyonu

$$P(Y=1 | \mathbf{z}, T) = \{1 + \exp[-\beta_0 - \beta' \mathbf{z}]\}^{-1} \quad (5)$$

biçiminde verilmektedir. T uzunluğundaki izleme periyoduna karşılık gelen yaşam olasılığı

$$P(Y=0 | \mathbf{z}, T) = S(T | \mathbf{z}, \beta) = \exp[-\beta_0 - \beta' \mathbf{z}] \{1 + \exp[-\beta_0 - \beta' \mathbf{z}]\}^{-1} \quad (6)$$

olmaktadır. (5) ve (6) eşitliklerinde izleme süresi T, bütün bireyler için aynı olmayı gerektirmektedir.

Cox (1972) tarafından önerilen

$$h(t, \mathbf{z}) = h_0(t) \exp(\alpha' \mathbf{z}) \quad (7)$$

modeli pozitif t zamanındaki hazarda dayalıdır. Bu modeldeki  $\alpha$  katsayı vektörü, (5) ve (6) eşitliklerindeki  $\beta_0$  ve  $\beta$  katsayılarıyla aynı olmayabilmektedir. (7) eşitliği ile verilen hazard, orantılı (proportional) olarak adlandırılmaktadır. Çünkü hazard, temel hazard olan  $h_0(t)$ 'nin ve hastalığın risk faktörlerinin bir fonksiyonu olan  $\exp(\alpha' \mathbf{z})$ 'nin çarpımı olarak ifade edilmektedir. Hazard modeline karşılık gelen yaşam (survival) fonksiyonu,

$$S(T | \mathbf{z}, \alpha) = \exp\left[-\int_0^T h_0(t) \exp(\alpha' \mathbf{z}) dt\right] \quad (8)$$

biçiminde verilmektedir. İzleme süresi T, Cox modelde herbir birey için farklı olabilmektedir.

Lojistik regresyon ile Cox modelin,  $h_0(t) = h$  özel durumu arasındaki ilişki belirtilmiştir. Lojistik model ve  $S(T | \mathbf{z}, \beta)$ 'nin (6) eşitliğinde verilen formu

$$\begin{aligned} -\ln[S(T | \mathbf{z}, \beta)] &= \ln\{[1 + \exp(-\beta_0 - \beta' \mathbf{z})][\exp(-\beta_0 - \beta' \mathbf{z})]^{-1}\} \\ &= \ln\left(\frac{1+u}{u}\right) = \ln\left(1 + \frac{1}{u}\right) \end{aligned}$$

biçiminde yazılabilmektedir. Burada  $u = \exp[-\beta_0 - \beta' \mathbf{z}]$  olmaktadır. Eğer  $u^{-1}$  küçük ise Taylor serisi açılımı ile

$$\ln\left(\frac{1+u}{u}\right) \cong \frac{1}{u}$$

yazılabilmektedir.

$$-\ln[S(T | \mathbf{z}, \boldsymbol{\beta})] \cong \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z}) \quad (9)$$

yaklaşımı,  $\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z})$ 'nin küçük olmasını sağlamaktadır. Böylece, (9) ifadesi ile verilen lojistik yaşam fonksiyonu ve (8) ifadesi ile verilen Cox model yaşam fonksiyonu

$$\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z}) \cong \int_0^T h_0(t) \exp(\boldsymbol{\alpha}'\mathbf{z}) dt$$

olduğunda yaklaşık olarak denk olmaktadır. Eğer temel hazard sabit ise ( $h_0(t) = h$ )

$$\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z}) \cong hT \exp(\boldsymbol{\alpha}'\mathbf{z}) \quad (10)$$

olmaktadır. (10) ifadesi ile verilen yaklaşık ilişki,  $\beta_0 = \ln(hT)$  ve  $i = 1, 2, \dots, r$  için

$$\beta_i \cong \alpha_i$$

olacaktır. İzleme uzunluğu  $T$ , lojistik regresyondaki  $\beta_0$  katsayısını doğrudan etkilemektedir. Lojistik regresyondaki risk faktörlerinin katsayıları, sabit temel hazard ve risk faktörlerinin orantılı log-doğrusal fonksiyonu ile Cox modelin katsayılarına yaklaşık olarak eşit olmaktadır. Bu sonuç,  $\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{z})$  küçük olduğu zaman sağlanmaktadır. Bu yaklaşım, olayın meydana gelmesi seyrek olduğunda ve risk faktörlerinin etkileri çok büyük olmadığı kısa izleme periyotları ile daha da iyileşebilmektedir. Diğer taraftan, daha genel olaylar için, uzun izleme periyotları durumunda ya da risk faktörlü olaylar için bu yaklaşım kötüleşmektedir (Green ve Symons, 1983). Yani, uzun izleme süresi durumunda Cox regresyon modeli, lojistik modelden daha çok bilgi vermekte ve ilgilenilen olay ve risk faktörleri arasındaki ilişkinin daha güçlü değerlendirmesini sağlamaktadır (Ingram ve Kleinman, 1989). Ayrıca, Cox regresyondan elde edilen katsayıların standart hataları azalmaktadır (Green ve Symons, 1983).

Lojistik regresyon modeli ile Cox regresyon modeli arasında iki ana fark vardır: Birincisi, her ikisi de farklı metrik kullanarak birlikteliği (ilişkiyi) ölçmektedirler. Cox regresyon süreyi dikkate alarak görel riski ölçerken, lojistik regresyon ise sabit izleme süresi varsayımı altında görel odds'u ölçmektedir. Bu durumda, olayın ortaya çıkma hızı yüksek olduğunda görel odds, görel riskten çok farklı olabilmektedir. İkinci fark ise, Cox model farklı izleme sürelerine izin vermektedir. Örneğin, durdurulmuş (censored) (bazı olayların ortaya çıkma zamanlarının bilinmediği fakat sadece gözlem periyodundan daha büyük olduğunun bilindiği durum) veriler için Cox regresyon modeli uygun olmaktadır. Lojistik model ise ikili (binary, dichotomous) (iki değer alıyor, örneğin, hastalık var-yok, yaşıyor-öldü gibi) olaylarla ilgili olarak bütün bireyler için izleme süresinin sabit olduğunu varsaymaktadır (Ingram ve Kleinman, 2001). Ancak bu

epidemiyolojik çalışmalarda çok nadir bir durumdur. Genellikle bireyler farklı gözlem periyotlarına sahip olmaktadır (Estévez vd, 2001).

Lojistik regresyonda zaman sabit alındığından izleme periyodunun başlangıcına yakın meydana gelen olayla, izleme periyodunun sonuna yakın meydana gelen olaya aynı ağırlık verilmektedir. Oysaki, gerçekte kabul edilen risk faktörünün bir hastalık riskine katkısının önemliliği, izleme periyodunun uzunluğuna bağlı olabilmektedir (Green ve Symons, 1983). Cox (1972), süreyi ve herbir birey üzerinden ölçülen açıklayıcı değişkenleri (covariates) de içeren bir model önermiştir. Cox modeli, tam başarısızlık zamanını kullanmakta ve durdurmayı da dikkate almaktadır (Chevret, 2001).

Bu farklılara rağmen, olayın ortaya çıkması seyrek olduğunda, risk faktörlerinin etkileri zayıf olduğunda ve izleme periyodu kısa olduğunda iki modelden elde edilen parametre tahminleri benzer olabilmektedir (Green ve Symons, 1983; Hauck, 1985; Ingram ve Kleinman, 1989; Chevret, 2001; Estévez, 2001). İzlemenin erken safhalarında iki model hemen hemen aynı sonucu vermektedir. Ancak, daha uzun izleme süreleri için iki modelden elde edilen regresyon katsayıları farklılık göstermektedir ve Cox regresyondan elde edilen katsayıların standart hataları azalmaktadır (Green ve Symons, 1983).

Lojistik model standart bir yaklaşım olarak görülürken son zamanlarda ki yeni istatistiksel yaklaşımlar Cox modelin kullanımını yaygınlaştırmışlardır. Epidemiyolojik çalışmalarda hastalar genellikle ölene kadar ya da takip edilebildiği kadar izlenmektedir. Bu da hastalar arasında gözlem periyotlarında büyük farklar yaratmaktadır. Lojistik modelde bu süre sabit varsayıldığından, Cox taraftarları, gözlem farklılıklarının lojistik modelde kullanılmadığını ifade etmişlerdir (Chevret, 2001). Cox regresyondaki katsayıların üstelliği ( $e^{\beta}$ ) hazard oranlarına (hazard ratios) karşılık gelmektedir. Bu da risk skorunu belirtmekte ve yüksek skorlar, olayın ortaya çıkması için yüksek riskleri göstermektedir (Estevez vd, 2001).

#### 4. MİDE KANSERİ İLE İLGİLİ BİR UYGULAMA

Lojistik ve Cox regresyon modellerinin uygulanması ve karşılaştırılması için Ankara Onkoloji Hastanesi'nde mide kanseri tanısı konan 118 hastaya ait veriler kullanılmıştır. Açıklayıcı değişkenler olarak yaş, cinsiyet, ağırlık kaybı, anemi, tümörün midedeki lokalizasyonu, lenf nodu diseksiyonunun genişliği, hastalığın evresi, tümörün mide duvarındaki invazyon derinliği, tümörün venöz, lenfatik ve perinöral invazyon yapması, adjuvant kemoterapi alınmıştır. 118 hastanın 52'si (%44.1) ölmüş, 66'sı (%55.9) ise çalışmanın sonunda yaşamını sürdürmektedir. Mide kanserinden ölümler olay olarak alınmış, halen yaşayanlar ise durdurulmuş olarak kabul edilmiştir. Ortalama yaş  $56,7 \pm 11,34$  (29-84 yıl) olarak bulunmuştur. 3 yıllık yaşam olasılığı %42.47'dir. Ortanca yaşam süresi ise  $29 \pm 6$ 'dır.

Kullanılan değişkenler ve düzeyleri Tablo 1.'de verilmiştir.

Lojistik ve Cox regresyon modellerinde her bir değişken için ilk düzey referans kategorisi olarak alınmıştır. Bu ise değişken düzeylerini yorumlamada önemli olmaktadır. Bütün testler için p değeri 0.05'den küçük ise önemli olarak kabul edilmiştir.

Başarısızlığı etkileyen risk faktörlerini belirlemek için öncelikle tek değişkenli lojistik ve Cox regresyon modelleri incelenmiştir. Modellerdeki değişken için  $\beta$  parametresi, standart hatası, p-değeri,  $\beta$  parametresinin risk oranı ( $\exp(\beta)$ ) ile önemli bulunan değişken düzeyleri için risk oranının alt ve üst sınırları verilmiştir.  $\beta$  parametresinin pozitif değer olması bu düzeyin referans kategorisine göre daha fazla riskli olduğunu,  $\beta$  parametresinin negatif değer olması ise bu düzeyin referans kategorisine göre daha az riskli olduğunu göstermektedir. Lojistik regresyonda odds oranı, Cox regresyonda ise risk oranı olarak ifade edilen  $\exp(\beta)$  değeri ise önemli bulunan düzeyin, referans kategorisine göre kaç kat (ya da % ne kadar) daha riskli olduğu yorumunu getirmektedir. Lojistik regresyondan elde edilen sonuçlar Tablo 2.'de, Cox regresyondan elde edilen sonuçlar ise Tablo 3.'de verilmiştir.

**Tablo 1.** Açıklayıcı Değişkenler ve Düzeyleri

| Değişken   | Değişken Düzeyleri                | n   | %    |
|--|-----------------------------------|-----|------|
| Cinsiyet (CİNS)  | 1. Erkek                          | 69  | 58.5 |
|  | 2. Kadın                          | 49  | 41.5 |
| Kilo Kaybı (KİLO)                                      | 1. Yok                            | 90  | 76.3 |
|  | 2. Var                            | 28  | 23.7 |
| Anemi (ANEMİ)  | 1. Yok                            | 31  | 26.3 |
|  | 2. Var                            | 87  | 73.7 |
| Lenf nodu diseksiyonunun genişliği (DİSEKSİYON)        | 1. D <sub>0-1</sub>               | 62  | 52.5 |
|  | 2. D <sub>2-3</sub>               | 56  | 47.5 |
| Tümörün midedeki lokalizasyonu (LOKAL)                 | 1. Üst üçlük                      | 21  | 17.8 |
|  | 2. Orta üçlük                     | 25  | 21.2 |
|  | 3. Alt üçlük                      | 62  | 52.5 |
|  | 4. Tüm mide                       | 10  | 8.5  |
| Tümörün mide duvarındaki invazyon derinliği (DERİNLİK) | 1. T <sub>1</sub> +T <sub>2</sub> | 17  | 14.4 |
|  | 2. T <sub>3</sub>                 | 57  | 48.3 |
|  | 3. T <sub>4</sub>                 | 44  | 37.3 |
| Hastalığın evresi (EVRE)                               | 1. Evre1+Evre2                    | 30  | 25.4 |
|  | 2. Evre3                          | 65  | 55.1 |
|  | 3. Evre4                          | 23  | 19.5 |
| Tümörün venöz invazyon yapması (VENÖZ)                 | 1. Yok                            | 87  | 73.7 |
|  | 2. Var                            | 31  | 26.3 |
| Tümörün lenfatik invazyon yapması (LENFATİK)           | 1. Yok                            | 84  | 71.2 |
|  | 2. Var                            | 34  | 28.8 |
| Tümörün perinöral invazyon yapması (PRİNÖRAL)          | 1. Yok                            | 102 | 86.4 |
|  | 2. Var                            | 16  | 13.6 |
| Adjuvant kemoterapi (KEMOTERAPİ)                       | 1. Yok                            | 34  | 28.8 |
|  | 2. Var                            | 84  | 71.2 |

Tablo 2. Tek Değişkenli Lojistik Regresyon Sonuçları

| Değişken   | $\beta$ | Standart Hatası | p     | Exp( $\beta$ ) | Alt sınır-Üst sınır |
|------------|---------|-----------------|-------|----------------|---------------------|
| YAŞ        | -.028   | .0169           | .0958 | .972           | -                   |
| CİNS       | -.369   | .379            | .3300 | .691           | -                   |
| KİLO       | .313    | .434            | .4699 | 1.368          | -                   |
| ANEMİ      | .672    | .440            | .1263 | 1.959          | -                   |
| DİSEKSİYON | -1.558  | .404            | .0001 | .210           | .095 - .465         |
| LOKAL      |         |                 | .0235 |                |                     |
| ORTA ÜÇLÜK | 1.098   | .617            | .0751 | 3.000          | -                   |
| ALT ÜÇLÜK  | .0953   | .533            | .8582 | 1.100          | -                   |
| TÜM MİDE   | 2.079   | .916            | .0232 | 7.999          | 1.328 - 48.182      |
| DERİNLİK   |         |                 | .0375 |                |                     |
| T3         | 1.978   | .797            | .0131 | 7.227          | 1.514 - 34.499      |
| T4         | 2.013   | .810            | .0130 | 7.485          | 1.529 - 36.645      |
| EVRE       |         |                 | .0018 |                |                     |
| EVRE3      | 2.087   | .592            | .0004 | 8.062          | 2.526 - 25.727      |
| EVRE4      | 1.958   | .680            | .0040 | 7.085          | 1.868 - 26.866      |
| VENÖZ      | 1.977   | .487            | .0000 | 7.224          | 2.781 - 18.765      |
| LENFATİK   | 1.377   | .432            | .0014 | 3.965          | 1.699 - 9.256       |
| PERİNÖRAL  | .867    | .554            | .1175 | 2.381          | -                   |
| KEMOTERAPİ | -.842   | .415            | .0422 | .431           | .191 - .971         |

Tablo 2'deki p değerleri incelendiğinde DİSEKSİYON, LOKAL, DERİNLİK, EVRE, VENÖZ, LENFATİK ve KEMOTERAPİ değişkenlerinin mide kanserinden başarısızlık (ölüm) için önemli risk faktörleri olduğu söylenebilmektedir ( $p < 0.05$ ). Önemli bulunan bu değişkenlerin her bir düzeyine karşılık gelen p değerlerine bakılarak önemli değişken düzeyleri belirlenebilmektedir. Her bir değişken için ilk düzeyler referans kategorisi olarak alındığından çizelgede yer almamaktadır. DİSEKSİYON değişkeni önemli bulunmuştur. Buna karşılık gelen katsayı da negatif olduğundan, lenf nodu diseksiyonunun  $D_{2-3}$  olmasının  $D_{0-1}$  olmasına göre başarısızlık riskinin %21 ( $\exp(\beta) = .21$ ) daha az olduğu söylenebilmektedir. LOKAL değişkeninin ise dördüncü düzeyi olan TÜM MİDE önemli olarak bulunmuştur. Buradan tümörün midedeki lokalizasyonu tüm mide olduğunda üst üçlük olmasına göre 8 kat ( $\exp(\beta) = 7.999$ ) daha fazla riskli olduğu yorumlanabilmektedir. Diğer değişkenler ve düzeyleri içinde aynı yorumlar yapılabilmektedir.

**Tablo 3.** Tek Değişkenli Cox Regresyon Sonuçları

| Değişken   | $\beta$ | Standart Hatası | p     | Exp( $\beta$ ) | Alt sınır-Üst sınır |
|------------|---------|-----------------|-------|----------------|---------------------|
| YAŞ        | -.023   | .013            | .0671 | .977           | -                   |
| CİNS       | -.304   | .288            | .2924 | .738           | -                   |
| KİLO       | .162    | .313            | .6062 | 1.175          | -                   |
| ANEMİ      | .585    | .353            | .0979 | 1.795          | -                   |
| DİSEKSİYON | -1.041  | .313            | .0009 | .3532          | .191 - .653         |
| LOKAL      |         |                 | .0307 |                |                     |
| ORTA ÜÇLÜK | .695    | .459            | .1302 | 2.004          | -                   |
| ALT ÜÇLÜK  | .199    | .434            | .6470 | 1.220          | -                   |
| TÜM MİDE   | 1.269   | .521            | .0149 | 3.558          | 1.281 - 9.885       |
| DERİNLİK   |         |                 | .0301 |                |                     |
| T3         | 1.619   | .732            | .0270 | 5.050          | 1.202 - 21.213      |
| T4         | 1.931   | .742            | .0093 | 6.894          | 1.610 - 29.514      |
| EVRE       |         |                 | .0007 |                |                     |
| EVRE3      | 1.952   | .534            | .0003 | 7.045          | 2.471 - 20.084      |
| EVRE4      | 2.154   | .595            | .0003 | 8.617          | 2.684 - 27.669      |
| VENÖZ      | 1.252   | .280            | .0000 | 3.498          | 2.021 - 6.055       |
| LENFATİK   | .842    | .283            | .0030 | 2.321          | 1.331 - 4.046       |
| PERİNÖRAL  | .867    | .554            | .1175 | 2.381          | -                   |
| KEMOTERAPİ | -.451   | .285            | .114  | .636           | .364 - 1.115        |

Tek değişkenli Cox regresyonu sonucunda elde edilen ve Tablo 3’de verilen p değerleri incelendiğinde DİSEKSİYON, LOKAL, DERİNLİK, EVRE, VENÖZ ve LENFATİK değişkenlerinin mide kanserinden başarısızlık için önemli risk faktörleri olduğu söylenebilmektedir ( $p < 0.05$ ). DİSEKSİYON değişkeni önemli bulunduğundan lenf nodu diseksiyonunun D<sub>2-3</sub> olmasının D<sub>0-1</sub> olmasına göre başarısızlık riskinin %35 ( $\exp(\beta) = .3532$ ) daha az olduğu söylenebilmektedir. LOKAL değişkeninin ise dördüncü düzeyi olan TÜM MİDE önemli olarak bulunmuştur. Buradan tümörün midedeki lokalizasyonu tüm mide olduğunda üst üçlük olmasına göre 3.5 kat ( $\exp(\beta) = 3.558$ ) daha fazla riskli olduğu yorumlanabilmektedir. Diğer değişkenler ve düzeyleri içinde aynı yorumlar yapılabilmektedir.

Tek değişkenli lojistik regresyondan elde edilen sonuçlarla tek değişkenli Cox regresyondan elde edilen sonuçları karşılaştırdığımızda KEMOTERAPİ değişkeni dışında aynı değişkenlerin önemli bulunduğunu ancak Cox regresyondan elde edilen katsayıların standart hatalarının daha küçük olduğu dikkat çekmektedir. Bu da lojistik regresyonla Cox regresyonu ayıran önemli bir özellik olmaktadır.

Bundan sonra ise tüm değişkenler birlikte ele alınarak adımsal seçim yöntemiyle lojistik ve Cox regresyon çözümlenmeleri yapılmış ve son adımda elde edilen sonuçlar sırasıyla Tablo 4. ve Tablo 5.’de verilmiştir.

**Tablo 4.**Adımsal Lojistik Regresyon Sonuçları

| Değişken   | $\beta$ | Standart Hatası | p     | Exp( $\beta$ ) | Alt sınır-Üst sınır |
|------------|---------|-----------------|-------|----------------|---------------------|
| DİSEKSİYON | -2.497  | .658            | .0001 | .082           | .023 - .298         |
| LOKAL      |         |                 | .0061 |                |                     |
| ORTA ÜÇLÜK | 3.698   | 1.138           | .0012 | 40.377         | 4.337 - 375.872     |
| ALT ÜÇLÜK  | 1.794   | .906            | .0477 | 6.012          | 1.018 - 35.498      |
| TÜM MİDE   | 3.553   | 1.433           | .0132 | 34.918         | 2.103 - 579.801     |
| EVRE       |         |                 | .0016 |                |                     |
| EVRE3      | 3.093   | .889            | .0005 | 22.047         | 3.859 - 125.938     |
| EVRE4      | 3.465   | 1.071           | .0012 | 31.965         | 3.917 - 260.819     |
| VENÖZ      | 2.791   | .757            | .0002 | 16.293         | 3.694 - 71.858      |
| KEMOTERAPİ | -3.304  | .873            | .0002 | .036           | .007 - .203         |

**Tablo 5.**Adımsal Cox Regresyon Sonuçları

| Değişken   | $\beta$ | Standart Hatası | P     | Exp( $\beta$ ) | Alt sınır-Üst sınır |
|------------|---------|-----------------|-------|----------------|---------------------|
| DİSEKSİYON | -.821   | .333            | .0137 | .440           | .229 - .845         |
| LOKAL      |         |                 | .0552 |                |                     |
| ORTA ÜÇLÜK | 1.168   | .490            | .0172 | 3.216          | 1.229 - 8.409       |
| ALT ÜÇLÜK  | .883    | .457            | .0536 | 2.417          | -                   |
| TÜM MİDE   | 1.390   | .546            | .0109 | 4.016          | 1.377 - 11.712      |
| EVRE       |         |                 | .0012 |                |                     |
| EVRE3      | 1.641   | .561            | .0034 | 5.162          | 1.718 - 15.505      |
| EVRE4      | 2.322   | .635            | .0003 | 10.200         | 2.937 - 35.425      |
| VENÖZ      | 1.016   | .307            | .0009 | 2.762          | 1.513 - 5.044       |
| KEMOTERAPİ | -1.190  | .317            | .0002 | .304           | .163 - .567         |

Adımsal seçim yöntemiyle elde edilen sonuçları karşılaştırdığımızda, her iki çözümlenin son adımında DİSEKSİYON, LOKAL, EVRE, VENÖZ ve KEMOTERAPİ değişkenlerinin olduğu modele ulaşılmış ve bu değişkenler önemli olarak bulunmuştur. Ancak lojistik regresyonda LOKAL değişkeninin ALT ÜÇLÜK düzeyi önemli olarak bulunurken Cox regresyonda bu düzey anlamlı olarak bulunmamıştır.

Elde edilen sonuçlardan Cox regresyonunun lojistik regresyona göre daha tutucu olduğunu söyleyebiliriz. Ayrıca Cox regresyondan elde edilen katsayıların standart hatalarının lojistik regresyondan elde edilen katsayıların standart hatalarına göre daha küçük olduğunu ve Cox regresyonu sonucunda elde edilen değerlerin daha güvenilir olduğunu söyleyebilmekteyiz.

## 5. SONUÇ

Bu çalışmada, lojistik ve Cox regresyon modelleri hakkında bilgi verildikten sonra aralarındaki matematiksel ilişkiden bahsedilmiş ve bu iki model karşılaştırılmıştır. Mide kanseri verileri kullanılarak da bir uygulama yapılmıştır. Bu modellerle başarısızlığı etkileyen önemli açıklayıcı değişkenler belirlenebilmekte, açıklayıcı değişkenlerin düzeyleri karşılaştırılabilmekte ve başarısızlık riskinin hangi düzeyde daha fazla olduğu yorumu yapılabilmektedir.

Olayın ortaya çıkması seyrek olduğunda, risk faktörlerinin etkileri zayıf olduğunda ve izleme periyodu kısa olduğunda lojistik ve Cox regresyon modellerinden elde edilen parametre tahminlerinin benzer olduğu, ancak daha uzun izleme süreleri için iki modelden elde edilen regresyon katsayılarının farklılık gösterdiği ve Cox regresyondan elde edilen katsayıların standart hatalarının azaldığı söylenebilmektedir.

Uygulamada, mide kanseri üzerinde etkili olduğu düşünülen yaş, cinsiyet, ağırlık kaybı, anemi, tümörün midedeki lokalizasyonu, lenf nodu diseksiyonunun genişliği, hastalığın evresi, tümörün mide duvarındaki invazyon derinliği, tümörün venöz, lenfatik ve perinöral invazyon yapması, adjuvant kemoterapi değişkenleri kullanılmıştır.

Başarısızlık üzerinde etkili olan değişkenleri belirlemek için yapılan tek değişkenli lojistik ve Cox regresyon modelleri sonucunda KEMOTERAPİ değişkeni dışında aynı değişkenlerin (DİSEKSİYON, LOKAL, DERİNLİK, EVRE, VENÖZ ve LENFATİK) ve değişken düzeylerinin önemli olarak bulunduğu, ancak Cox regresyondan elde edilen katsayıların standart hatalarının daha küçük olarak elde edildiği görülmüştür.

Lojistik ve Cox regresyonda tüm değişkenler birlikte ele alınarak adımsal seçim yöntemi uygulandığında her iki modelin son adımında aynı değişkenlerin (DİSEKSİYON, LOKAL, EVRE, VENÖZ ve KEMOTERAPİ) olduğu modele ulaşıldığı ve bu değişkenlerin önemli olarak bulunduğu görülmüştür. Sadece lojistik regresyonda LOKAL değişkeninin ALT ÜÇLÜK düzeyi anlamlı bulunurken Cox regresyonda bu düzey anlamlı olarak bulunmamıştır. Ayrıca Cox regresyondan elde edilen katsayıların standart hataları lojistik regresyondan elde edilen katsayıların standart hatalarına göre daha küçük elde edilmiştir.

## KAYNAKLAR

- CHEVRET, S. (2001), *Logistic or Cox Model to Identify Risk Factors of Nosocomial Infection: Still a Controversial Issue*, Intensive Care Med., 27, 1559-1560.
- COX, D.R. (1972), *Regression Models and Life Tables*, Journal of Royal Statistical Society, Series B, 34, 187-202.
- COX, D.R. and OAKES, D. (1984), *Analysis of Survival Data*, Chapman and Hall.

- ESTÉVEZ, J.I., CONCHA, D.M., MOLINA, C.D., CALLES, J.M., CASTILLO, A.S. AND NAVAJAS, R.F.C. (2001), *Comparison of Different Methodological Approaches to Identify Risk Factors of Nosocomial Infection in Intensive Care Units*, *Intensive Care Med.*, 27, 1254-1262.
- GREEN, M.S. and SYMONS, M.J. (1983), *A Comparison of the Logistic Risk Function and the Proportional Hazards Model in Prospective Epidemiologic Studies*, *J. Chron. Dis.*, 36, 10, 715-724.
- HAUCK, W.W. (1985), *A Comparison of the Logistic Risk Function and the Proportional Hazards Model in Prospective Epidemiologic Studies*, *J. Chron. Dis.*, 38, 1, 125-126.
- HOSMER, D.W. and LEMESHOW, S. (1999), *Applied Survival Analysis: Regression Modelling of Time to Event Data*, John Wiley and Sons.
- INGRAM, D.D. and KLEINMAN, J.C. (1989), *Empirical Comparisons of Proportional Hazards and Logistic Regression Models*, *Statistics in Medicine*, 8, 525-538.
- KALBFLEISCH, J. D. and PRENTICE, R.L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons.
- LAWLESS, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons.
- SCHEMPER, M. (1992), *Cox Analysis of Survival Data with Nonproportional Hazards Functions*, *The Statistician*, 41, 455-465.
- ÖZDAMAR, K. (1997), *Paket Programlar ile İstatistiksel Veri Analizi I*, Anadolu Üniversitesi.

## Review of Logistic and Cox Regression Models

### ABSTRACT

*In this study, Logistic regression and Cox regression models used to identify risk factors of any event in epidemiologic studies are investigated. These models and mathematical relationship and comparisons between the them are explained. An application is carried out by using gastric cancer data.*

**Key Words:** *Logistic Regression, Cox Regression, Survival Analysis*