

Yaşam Analizinde Cox Regresyon Modeli ve Artıkların İncelenmesi

Meral YAY¹, Elif ÇOKER¹, Ömer UYSAL³

¹ Mimar Sinan Güzel Sanatlar Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, İstanbul

² İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Biyoistatistik Bilim Dalı, İstanbul

Özet

Yaşam analizinde Cox regresyon modeli, yaşam süresi üzerinde etkili olan değişkenleri, faktörleri belirlemede sık kullanılan bir yöntemdir. Yöntem, başta tıp olmak üzere yaygın biçimde kullanılmaktadır. Bu çalışmada, yeni doğum yapmış 100 kadın üzerinde yapılan araştırmalardan elde edilen veriler kullanılmıştır. Bu kadınlar doğumdan hemen sonra başlayarak dört ay boyunca izlenmişlerdir. İzlemenin amacı emzirme sürelerini etkileyen faktörleri ortaya çıkartmaktır.

Anahtar Kelimeler: Cox regresyon modeli, yaşam analizi, Schoenfeld artıkları

Cerrahpaşa Tıp Derg 2007; 38: 139 - 145

Abstract

Cox regression model in survival analysis and examining the residuals

Cox regression model in survival analysis is a method widely used to determine the factors that are effective on survival time. Especially, this method is frequently used in medical research. In this study, research data are taken from 100 women which have just gave a birth. These women have been observed throughout the four months after their delivery. The aim of this process is to ascertain the factors that are effective on lactation time.

KeyWords: Cox regression model, survival analysis, Schoenfeld residuals

Cerrahpaşa J Med 2007; 38: 139 - 145

Yaşam analizi (izleme analizi) tıpta çeşitli hastalıklara ilişkin risk, prognostik etkenler (açıklayıcı değişkenler, bağımsız değişkenler, ortak değişkenler, belirleyici etkenler), tedavi başarısı vb. verilerin incelenmesinde büyük önem taşımakta ve tıbbın birçok alanında uygulanmaktadır. Yaşam analizinde sıkça kullanılan yöntemlerden biri de, yaşam süresi üzerinde etkili olan faktörlerin belirlenmesinde kullanılan Cox Regresyon modelidir. Model, ismini kendisini geliştiren David R. Cox (1972)'dan almıştır [1].

Yaşam analizi, belirli bir başlangıç noktasından başlayan (hastalığın başlangıcı, tedavinin başlangıcı) izleme sürecindeki olgunun sonuca varmasıdır. Analiz; ölüm, tedaviye yanıt almak gibi farklı durumlara sonuçlanabilir. İzleme sürecinde izlenen gözlemlerden bazılarının hastalık dışındaki nedenler yüzünden ölmesi veya araştırmayı yapan kişi tarafından izlenmekten vazgeçilmesi durumları söz konusu olabilir. Bu gibi durumlarda analizin sonuçlanmaması son derece olağandır. Bu tür gözlemlere tamamlanmamış (censored) gözlemler denir. Tamamlanmamış veriler gözlemlerin tamamının değil bir kısmının bulunduğu

veriler olarak da açıklanabilir [2].

Yaşam analizi, ilgilenilen olayın gerçekleşmesine kadar geçen süredeki verilerin analizinde kullanılır. Analizin bağımlı değişkeni, yukarıda sıralanan üç farklı durumdan birinin gerçekleşmesine kadar geçen süre olup, "başarısızlık süresi" veya "yaşam süresi" olarak adlandırılır [3]. Bağımlı değişken, kardiyovasküler ölüme kadar geçen süre, hamile kalıncaya kadar geçen süre vb. ile örneklenir. Olağan dışı bir durumla karşılaşmadıkça bağımlı değişkenin gözlenmemesi (tamamlanmaması) söz konusu olmaz. Bununla birlikte yaşam analizi, bazı gözlemler tamamlanmamış olsa da uygulanabilmektedir. Bu, yaşam analizinin güçlü yanlarından biri olarak kabul edilir.

Yaşam analizinin temel amaçları aşağıdaki gibi üç başlık altında toplanabilir [4]:

- Yaşam karakteristiklerini tahmin etmek ve yorumlamak için Kaplan Meier yöntemi ve grafikleri, medyan tahmini ve güven aralıkları,
- Farklı gruplardaki yaşam sürelerini karşılaştırmak için Logrank testi,
- Bağımsız değişkenler ile yaşam süresi arasındaki ilişkinin incelenmesi için Cox regresyon modeli.

Yaşam analizinin bağımsız değişkenleri normal dağılmazlar; ayrıca bu değişkenler arasında oransal ilişkiler söz konusudur yani değişkenler bağımsız değildir.

Alındığı Tarih: 5 Ekim 2007

Yazışma Adresi (Address): Dr. Ömer UYSAL

İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Temel Bilimler Biyoistatistik Bilim Dalı
34098 Cerrahpaşa - İstanbul

E-posta: omuyusal@istanbul.edu.tr

Değişkenlerin bu özelliklerinden dolayı yaşam verilerinin analizinde klasik çoklu regresyon değil, Cox regresyon yöntemi kullanılır.

YÖNTEM

Cox Regresyon Modeli

Cox regresyon yöntemi, bir hastanın yaşam süresi (bağımlı değişken) ile birden fazla bağımsız değişken arasındaki ilişkiyi ortaya çıkaran istatistiksel bir yöntemdir. Cox regresyon analizinde hedef, yaşam verilerinin genel durumunu yansıtacak bir model oluşturmaktır. Bu yolla, yaşam süresi üzerinde etkili olduğu düşünülen bağımsız değişkenlerin etkileri eşanlı olarak açıklanabilmekte, yani ölçülebilmektedir. Cox regresyon modeli, klinik bir çalışmada hastaların yaşamını analiz etmek için kullanıldığında, tedavinin etkisini diğer değişkenlerin etkisinden ayırmaya yardımcı olur.

Parametrik modellerin gerektirdiği varsayımların (normallik, bağımsızlık vb.) sağlanmadığı durumlarda Cox regresyon analizi, parametrik analizlerden daha etkilidir. Cox regresyon modelinin temel varsayımları şu şekilde açıklanabilir: (1) Bağımsız değişkenlerin risk (hazard) fonksiyonu üzerindeki etkileri loglineerdir. (2) Bağımsız değişkenlerin loglineer fonksiyonu ile risk fonksiyonu arasındaki ilişki çarpımsaldır [5]. Bu iki varsayıma ek olarak gözlemlerin birbirinden bağımsız olmaları ve risk oranının zamana göre değişmemesi, yani sabit olması gerekmektedir. Risk oranı ile ilgili bu varsayım oransal risk varsayımı olarak bilinir. Oransal risk varsayımı sağlanmadığında, zamana bağlı bağımsız değişkenli Cox regresyon modeline başvurulur. İncelenen örnekte bağımsız değişkenler yoksa veya sadece bir kategorik bağımsız değişken varsa risk fonksiyonları için yaşam tablolarına ve Kaplan Meier yöntemine başvurulabilir.

χ bağımsız değişkenler vektörü olsun. Yaşam süresi (T), birikimli dağılım fonksiyonu $P(t) = \Pr(T \leq t)$ ve olasılık yoğunluk fonksiyonu $P(t) = dP(t)/d(t)$ olan bir raslantı değişkeni olarak tanımlansın. Analiz için T yaşam süresi içinde herhangi bir t_i zamanında ölümle (başarısızlıkla) karşılaşacağı varsayılır. Yaşam analizinde her bir gözlem için başarısızlık olarak adlandırılan ve başarısızlık zamanı (failure time) olarak tanımlanan bir nokta olay söz konusudur. Buradan hareketle risk fonksiyonunun tümleyeni olan yaşam fonksiyonu $S(t)$ aşağıdaki eşitlikle açıklanır [6].

$$S(t) = P(T) = 1 - P(t) \quad (1)$$

Bir gözlemin χ bağımsız değişkenlerine ilişkin risk fonksiyonu $h(t, \chi)$ ile gösterilir ve aşağıdaki gibi ifade edilir.

$$h(t, \chi) = h_0(t) \exp(\beta \chi) \quad (2)$$

Burada $h_0(t)$, $\chi = 0$, olduğu durumdaki temel risk fonksiyonu, β ise bağımsız değişkenlerin regresyon katsayıları vektörüdür [7].

Yaşam süresi T'nin risk fonksiyonu $h(t, \chi)$, koşullu bir başarısızlık oranıdır. Bu fonksiyon, çok kısa bir zaman

aralığındaki başarısızlık olasılığı olarak da tanımlanabilir. Fonksiyonda, her bir bireyin aralığın başlangıcından itibaren gözlemlendiği varsayılır. Çok kısa bir zaman aralığı ($t, t + \Delta t$) olarak ifade edildiğinde risk fonksiyonu aşağıdaki gibidir:

$$h(t, x) = \frac{\lim_{\Delta t \rightarrow 0} P \left[\begin{array}{l} \text{bir bireyin } t \text{ zamanına kadar yaşadığı bilindiğine göre} \\ \text{bu bireyin } (t, t + \Delta t) \text{ zaman aralığına düşmesi} \end{array} \right]}{\Delta t} \quad (3)$$

Risk fonksiyonu, birikimli fonksiyon F(t) ve olasılık yoğunluk fonksiyonu f(t) cinsinden,

$$h(t, x) = \frac{f(t)}{1 - F(t)} \quad (4)$$

olarak da tanımlanabilir [8].

χ bağımsız değişkenlerine ilişkin vektörün bir tek değişken içermesi durumunda regresyon katsayıları,

$$h(t, x) = h_0(t) e^{(\beta x_1)} \quad (5)$$

birden fazla değişken varlığında ise,

$$h(t, x) = h_0(t) e^{(\beta x_1 + \beta x_2 + \dots + \beta x_p)} \quad (6)$$

ile tahmin edilir. Yukarıdaki eşitlikte yer alan $\chi_1, \chi_2, \dots, \chi_p$ bağımsız değişkenleri, T yaşam süresi üzerinde etkili olan kan basıncı, sıcaklık, yaş gibi sürekli değişkenler veya histolojik evreler, hastalık evresi gibi kategorik değişkenler olabilir.

β bilinmeyen katsayılar vektörünün tahmini için kısmi olabilirlik fonksiyonundan faydalanılır. Cox regresyon analizinde kullanılan kısmi olabilirlik fonksiyonu için öncelikle gözlenen n tane yaşam süresi içerisinde k tanesinin $t_1 < t_2 < t_3 < \dots < t_k$ şeklinde sıralanmış başarısızlık süreleri olduğunu düşünelim. i. olayın meydana gelme zamanında başarısızlığa uğrama riski altında bulunan bireylerin oluşturduğu kümeyi de risk kümesi olarak isimlendirelim ve $R(t_i)$ ile gösterelim [1]. Bütün bu tanımlar ve açıklamalar çerçevesinde Cox tarafından β 'nin tahmin edilmesinde kullanılmak üzere önerilen kısmi olabilirlik fonksiyonu aşağıda gösterilmiştir.

$$L(\beta) = \prod_{k=1}^n \frac{\exp(\beta' x_k)}{\sum_{j \in R(t_k)} \exp(\beta' x_j)} \quad (7)$$

Kısmi olabilirlik fonksiyonunun logaritmik formu aşağıdaki gibidir.

$$\log L(\beta) = \sum_{k=1}^n \left\{ \beta x_k - \log \left[\sum_{j \in R(t_k)} \exp(\beta x_j) \right] \right\} \quad (8)$$

β katsayılarının en çok olabilirlik tahminleri, yukarıdaki logaritmik kısmi olabilirlik fonksiyonunun en büyüklenmesiyle elde edilir. Herhangi bir bağımsız değişken için elde edilen pozitif bir katsayı riskin yüksek ve prognostik etkilerin kötü olduğunu işaret eder. Negatif bir katsayı ise,

gözlemler için daha iyi prognostik etkenlerin olduğu şeklinde yorumlanır [9].

Modeldeki β Katsayılarının Anlamlılık Testleri

β katsayılarının anlamlılık testleri için sıfır hipotezi $H_0: \beta = 0$ şeklinde ifade edilir. Cox modelindeki β katsayılarının anlamlılığını test etmek amacıyla kullanılan üç yöntem vardır. Bunlar:

- Wald testi
- Olabilirlik oran testi

ve

- Skor testidir.

β 'ların anlamlılığını test etmede kullanılan Wald test istatistiği, t veya z istatistiğinin genelleştirilmiş şekli olup, en çok olabilirlik tahmini (MLE) ve hipotezlenen değer arasındaki farkın bir fonksiyonudur. Bu fark, en çok olabilirlik tahmininin standart hatasının tahmini değeri ile normalleştirilmiştir. Tek bir değişken için test istatistiği,

$$W = (\beta / se_{\beta})^2 \quad (9)$$

olarak tanımlanır [7]. Burada n 'nin büyük değerleri için, 1 serbestlik dereceli W dağılımı gösterir. Karar için, χ^2 hesaplanan W değerinin, χ^2 tablo değeri ile karşılaştırılması yeterlidir.

Olabilirlik oran testi, hipotezlenen parametre değerleri içinde maksimum olan değer in olabilirliğinin oranına dayanır. Test istatistiği (LR) şöyledir:

$$LR = -2 (\log (l_0) / \log (l_v)) \quad (10)$$

Burada l_0 , tam modelin maksimum olabilirlik istatistiğini, l_v ise v değişkenli modelin maksimum olabilirlik istatistiğini ifade etmektedir. Çok sayıda örnek için LR istatistiği, tahmin edilecek parametre sayısına eşit serbestlik derecesi ile χ^2 (burada χ^2_v) dağılımı gösterir. Burada da karar için hesaplanan LR değerinin, χ^2 tablo değeri ile karşılaştırılması yeterlidir.

Bu üç test içinde özel bir yeri olan skor testi ile iki grubu karşılaştırmada kullanılan Mantel-Haenszel Log-rank testi sonuçları benzerdir. Olabilirlik oran testinin, skor testinden daha iyi sonuçlar verdiği durumlarla da karşılaşılabilir. Skor testi, birden fazla değişken ve sürekli değişkenler için log rank testinin genelleştirilmiş şeklidir. İki değerli değişkenin gözlenme oranı "p" olsun. Bu orana göre ilgilenilen değişkenin r kez gözlenme olasılığını kullanarak maksimum olabilirlik fonksiyonu aşağıdaki gibi yazılır:

$$L = \log(l) = r \log p + (n - r) \log (1 - p) \quad (11)$$

Skor test istatistiğini elde etmek için, L 'nin 'p' ye göre alınan birinci ve ikinci türevleri sıfıra eşitlenerek çözülür ve aşağıdaki oran elde edilir:

$$S = (L')_2 / (L'') \quad (12)$$

Skor test istatistiği, serbestlik derecesi tahmin edilen parametre sayısı olmak üzere χ^2 dağılımına uyar [5]. S için hesaplanan değer, χ^2 tablo değeri ile karşılaştırılarak istatistiğin anlamlılığı test edilir.

Artık Analizi- Schoenfeld Artıkları

Cox regresyon modeli için dört çeşit artık tanımlanmaktadır. Bu bağlamda dört temel artık analizinden söz edilir. Bunlar; Cox-Snell, martingale, sapma (deviance) ve Schoenfeld (skor) artıklarıdır. Bu çalışmanın uygulamasında Schoenfeld artık analizi kullanıldığından, burada sadece anılan artık analizinin açıklanmasıyla yetinilecektir.

Cox-Snell, martingale ve sapma artıklarının iki önemli dezavantajı söz konusudur. Bu dezavantajlar, artıkların ağırlıklı olarak gözlenen yaşam süresine bağlı olmalarını ve birikimli risk fonksiyonunun tahminini gerektirmesidir. Schoenfeld tarafından önerilen, skor artıkları olarak da adlandırılan Schoenfeld artıklarında bu sorunlar giderilmiştir. Bu yönüyle Schoenfeld artıkları, diğer artıklardan önemli bir farklılık göstermektedir. Bu artıklarda, her bireyin artığı için tek bir değer yerine, tahmin edilmiş olan Cox regresyon modelinde yer alan her bir bağımsız değişken için birer tane olmak üzere değerler kümesi yer almaktadır.

Model içeriğindeki j.bağımsız değişken X_j 'nin i. Schoenfeld artığı aşağıdaki formül yardımıyla hesaplanır:

$$r_{s_{ji}} = \delta_i \{ x_{ji} - a_{ji} \} \quad (13)$$

Burada x_{ji} , $j = 1, 2, \dots, p$ olmak üzere çalışmadaki i. birey için j. bağımsız değişkenin değeridir.

$$a_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\hat{\beta}' x_l)}{\sum_{l \in R(t_i)} \exp(\hat{\beta}' x_l)} \quad (14)$$

a_{ji} , ağırlıklı risk ortalaması olarak da anılır. Formüldeki $R(t_i)$, t_i zamanında risk altındaki tüm bireylerin oluşturduğu gruptur. Bu artıklar için sıfırdan farklı değerlerin yalnızca, tamamlanmış gözlemler için ortaya çıkacağına dikkat edilmelidir. Schoenfeld artıkları kısaca, bağımsız değişkenin gerçek değeri ile ağırlıklı risk ortalaması arasındaki fark olarak tanımlanır.

Bu artığın elde edilmesi için birkaç yol vardır. Schoenfeld artığı, modeldeki j. parametre için etkin skorun i. bileşeninin bir tahminidir. Etkin skor, β_j 'ye ilişkin kısmi olabilirlik fonksiyonunun logaritmasının birinci türevine eşittir:

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^n \delta_i \left\{ x_{ji} - \frac{\sum_{l} x_{jl} \exp(\hat{\beta}' x_l)}{\sum_{l} \exp(\hat{\beta}' x_l)} \right\} \quad (15)$$

β_j 'nin tahminindeki i. terim, x_j için Schoenfeld artığını

oluşturmaktadır.

Oransal hazard varsayımı Schoenfeld artıkları ile kontrol edilebilir [10]. Oransal risk varsayımının geçerliliği zamana bağlı çizilen Schoenfeld artıkları grafiği ile kontrol edilebilir. Çizilen Schoenfeld artıkları grafiği, yatay bir doğru etrafında seyrediyorsa, oransal hazard varsayımının sağlandığı söylenecektir [11].

t_i zamanında risk grubunda bulunan l . birey için kullanılan ağırlık $\exp(\beta'x_i)$ 'dir ve (13) nolu eşitlikteki x_{ij} kat sayısı, bu birey için maksimize edilmiş kısmi olabilirlik fonksiyonunun bileşeni olan;

$$\frac{\exp(\hat{\beta}'x_i)}{\sum_{i \in R(t_i)} \exp(\hat{\beta}'x_i)}$$

'ye eşittir. β 'ların tahmini için; $\frac{\partial \log L(\beta)}{\partial \beta_j} \Big|_{\hat{\beta}} = 0$ olduğundan,

Schoenfeld artıklarının toplamı sıfır olmalıdır. Bu artıklar için, büyük örneklerde $r_{s_{ij}}$ 'nin beklenen değerinin sıfıra eşit olduğu ve artıkların bağımsız olduğu söylenebilir [12].

Bu artıkları kullanırken, t_i zamanında ölüm olasılıkları düşük olan bireyler, t_i zamanında ölüm riski altında olan bireylere göre, daha düşük değerli artıklara sahip olacaktırlar. Dolayısıyla, ölüm riski altında olan bireylere ilişkin artıklar, ölüm riski az olan bireylere ilişkin artıklara oranla daha büyük olacaktır.

BULGULAR

Bu çalışmada, 1998 yılında İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Çocuk Sağlığı ve Hastalıklar Anabilim Dalı Yenidoğan Ünitesinde izlenen 100 olgu üzerinde yapılan araştırmada elde edilen veriler kullanılmıştır [13]. Bu olgular doğumun hemen ardından başlayarak dört ay boyunca izlenmişlerdir. İzlemenin amacı annelerin emzirme sürelerini başka bir ifadeyle bebeklerin süttan kesilme noktasını etkileyen faktörleri ortaya çıkartmaktır. Elde edilen verilere Cox regresyon yöntemi uygulanmıştır.

İstatistiksel olarak anlamlı olan değişkenleri belirlemek amacıyla öncelikle "p" değerlerine bakılmış, yorumlar sadece anlamlı olan değişkenler üzerinden yapılmıştır. Cox regresyon analizinde elde edilen katsayıların üstel hale dönüştürülmesi sonucunda, ilgili değişkene ait göreceli değerler hesaplanır. Analizde kategorik değişken(ler) için, hangi değişken düzeyi en az etkili ise, o düzey referans kategorisi olarak alınır. Değişkene ilişkin düzey(lerin) yorumu da referans kategoriye göre yapılır. Bu bize kategorik değişkenlerde referans kategoriye göre, ilgili kategorinin sonucu kaç kat daha fazla ya da az değiştirdiğini ifade eder. β parametresinin pozitif değer olması, bu düzeyin referans kategorisine göre daha fazla riskli olduğunu, negatif değer olması ise bu düzeyin referans kategorisine göre daha az riskli olduğunu göstermektedir [14]. Nicel değişkenlerde ise değişkenin bir birim artması durumu, sonucu hangi oranda arttırdığı ya da azalttığı bilgisini ve-

rir. İstatistiksel olarak anlamlı bulunmayan değişkenlerde bu yorumlar yapılmaz.

Göreceli değerler, sınır kabul edilen 1'in üstünde ya da altında oluşlarına göre o değişkenin sonucu artırıcı ya da azaltıcı etkisini ortaya koyarlar. Tıpta bu durumlara sırasıyla etyolojik veya prevantiv isimleri verilir. Anlamlılık durumunda güven aralığı da sonucun etyolojik ya da prevantiv olması durumuna göre, alt ve üst sınırların her ikisi de 1'in üstünde ya da altında olacaktır [15]. Göreceli değeri 1'in üstünde kalan değişkenin anlamlı olduğu söylenecektir. Risk oranı (e^{β} , hazard oranı, OR= Odds Ratio), önemli bulunan düzeyin referans kategorisine göre kaç kat (ya da % ne kadar) daha riskli olduğu yorumunu getirmektedir. "p" değerlerine göre verilen anlamlılık kararını, risk oranı ve risk oranı için bulunan güven aralıklarında desteklemelidir.

Çalışmada, yeni doğum yapan kadınların doğumu takiben ilk dört aydaki emzirme süreleri "yaşam süresi" olarak alınmıştır. Bebeğin süt emmeyi kesmesi tamamlanmamış (censored) gözlem olarak değerlendirilmiştir. Bebeğin süt emmeyi kesmesi "1", kesilmemesi "0" olarak kodlanmıştır ve sonuç değişkeni olarak ele alınmıştır. Sonuç değişkenini etkileyen faktörler, yani bağımsız değişkenler sırasıyla; annenin yaşı (yıl), bebeğin doğum ağırlığı (gram), annenin bekleğinde anne sütü alıp almadığı (1/0), bebeğin cinsiyeti (kız = 1, erkek = 0), bebeğin doğumu ile anneden sütü aldığı ilk an arasında geçen zaman (saat) ve hastanede bebeğe mama verilip verilmediği (1/0) şeklinde belirlenmiştir.

Sonuç değişkenini etkileyen faktörlerin tümünün modele dahil edilmesiyle ulaşılan tam model bulguları Tablo 1'de özetlenmiştir.

Tablo 1'de açıklanan bulgular doğrultusunda, bebeğin süt emmeyi kesmesini istatistiksel olarak anlamlı biçimde etkileyen değişkenlerin; bebeğin doğum ağırlığı, ilk süt verme zamanı ve hastanede mama verilmesi dolayısıyla bebeğin mamayla tanışması olduğu görülmüştür. Bu değişkenlerin "p" değerleri sırasıyla 0.036, 0.008 ve 0.001'dir. Daha önce de ifade edildiği gibi anlamlılık kararı "p" değerlerine göre verilmektedir. Bu bağlamda $\alpha=0.05$ için anılan değişkenlerin anlamlı olduğu görülmüştür. Ayrıca her üç değişkenin e^{β} risk oranı 1'in üzerindedir. Bu sonuç "p" değerleri için elde edilen anlamlılık durumunu desteklemektedir. Risk oranı için belirlenen % 95 güven aralıkları incelendiğinde ise, doğum ağırlığı için (1.000 - 1.002), ilk süt verme zamanı için (1.015 - 1.107) ve firmanın eşantıyon mama vermesi için (1.620 - 4.500) olarak elde edilmiştir. "p" değerleri, e^{β} risk oranı ve % 95 güven düzeyine göre belirlenen güven aralıklarının uyum içinde ve birbirlerini destekler nitelikte olduğuna işaret etmektedir. Buna göre bebeğin doğum ağırlığı ne kadar yüksekse ve ilk süt verme saati doğumdan sonra ne kadar geç olursa, bebeğin süt emmeyi daha erken kesmesi olasılığı artmaktadır. Ayrıca hastanede bebeğin mama almış olması da, bebeğin süt emmeyi daha erken kesmesi olasılığını artırmaktadır.

Tablo 1. Tam modele ilişkin Cox Regresyon sonuçları.

Değişkenler	β	SE_{β}	Wald	sd	p	$e^{\beta} = OR$	$e^{\beta} = OR$ için % 95 güven aralığı	
							Alt	Üst
Hastanede bebeğe mama verilmesi	0.990	0.2600	14.54	1	<0.001	2.700	1.620	4.500
Annenin bebekliğinde anne sütü alması	-0.140	0.3100	0.19	1	0.660	0.870	0.470	1.610
Annenin yaşı	0.007	0.0300	0.06	1	0.810	1.007	0.950	1.070
Bebeğin cinsiyeti	0.148	0.1300	1.22	1	0.270	1.160	0.890	1.510
Bebeğin doğum ağırlığı	0.001	0.0001	4.42	1	0.036	1.001	1.000	1.002
İlk süt verme zamanı	0.060	0.0220	6.98	1	0.008	1.060	1.015	1.107

Tablo 1. İndirgenmiş modele ilişkin Cox Regresyon sonuçları.

Değişkenler	β	SE_{β}	Wald	sd	p	$e^{\beta} = OR$	$e^{\beta} = OR$ için % 95 güven aralığı	
							Alt	Üst
Hastanede bebeğe mama verilmesi	0.972	0.250	15.053	1	<0.001	2.640	1.620	4.320
Bebeğin doğum ağırlığı	0.001	0.000	4.327	1	0.038	1.001	1.000	1.001
İlk süt verme saati	0.061	0.022	7.936	1	0.005	1.063	1.019	1.109

Bilindiği gibi Tablo 1'deki sonuçlar tam modele aittir. Anılan tablodaki sonuç değerlerinin kullanılmasıyla sadece anlamlı bulunan değişkenler ile yeni bir model (indirgenmiş) elde edilmiştir. İndirgenmiş modele ilişkin sonuçlar Tablo 2'de verilmiştir.

Tablo 2 incelendiğinde, tam modele ilişkin açıklamaların yanısıra referans kategoriye göre OR oranlarını da yorumlamak mümkündür. Hatırlanacağı üzere indirgenmiş modeldeki tek kategorik değişken olan hastanede bebeğe mama verilmesi (1), verilmemesi ise (0) ile kategorize edilmişti. Analizde yorum referans kategoriye göre yapılmakta ve anılan değişken için (1) kategorisi göz önünde bulundurulmaktadır. Buna göre, bebeğe hastanede mama verilmesi, bebeğin süt emmeyi kesmesi riskini kesmemesine göre yaklaşık 2.7 kat arttırmaktadır. Bebeğe hastanede mama verilmesi kategorik değişkenine ilişkin çizilen yaşam fonksiyonu grafiği (Şekil 1), bebeğe hastanede mama verilmesi durumunun verilmemesine nazaran emzirme süresini azalttığını göstermektedir.

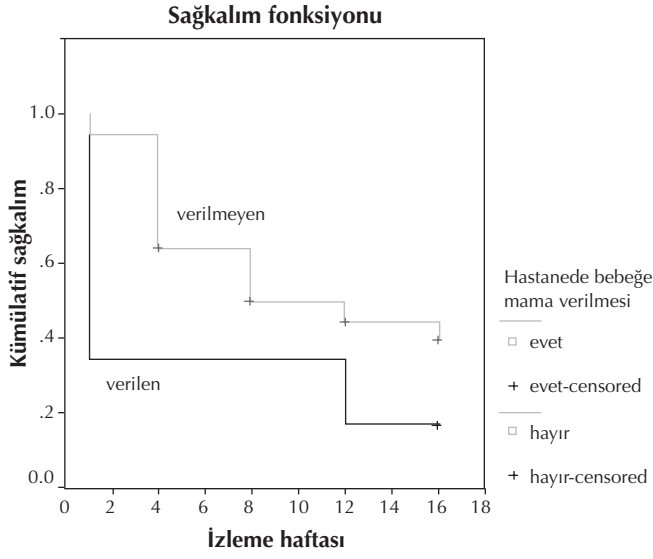
Bebeğe hastanede mama verilmesi kategorik değişkenine ilişkin çizilen risk fonksiyonu grafiğinde ise (Şekil 2), bebeğe hastanede mama verilmesi durumunda verilmemesine oranla emzirmenin erken kesilme riskinin arttığını görülmektedir.

Geriye doğru seçim tekniği uygulamasının 4. adımında

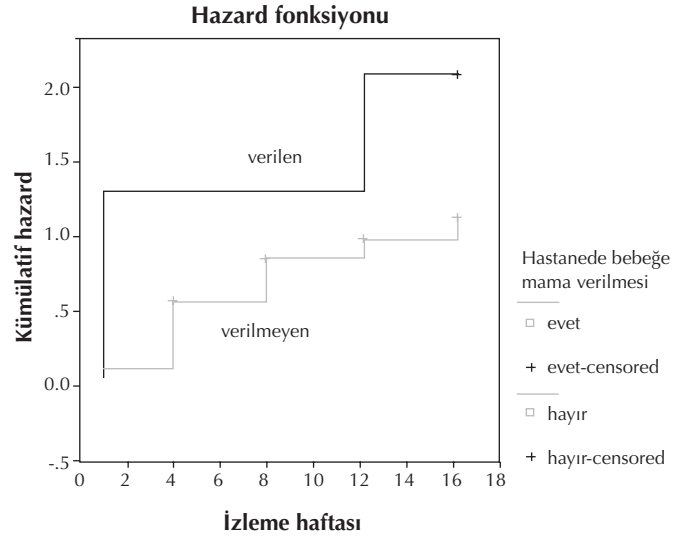
ulaşılan sonuçlara göre; firmanın mama hediye etmesi, bebeğin doğum ağırlığı ve ilk süt verme zamanı değişkenleri indirgenmiş en iyi modelin değişkenleridir. İndirgenmiş model ve tam modele ilişkin "p" değerleri, e^{β} risk oranı ve risk oranı için belirlenen % 95 güven aralıkları birbirleriyle tutarlılık göstermektedir. Aynı sonuçlar ileriye doğru seçim tekniği uygulamasının 3. adımında da elde edilmiştir.

Modelin nicel değişkenleri olan bebeğin doğum ağırlığı ve ilk süt verme zamanına ilişkin Schoenfeld artıkları grafiği çizildiğinde, yatay doğru etrafında toplanan ilk süt verme zamanı grafiğinin tam da istenen biçimde olduğu görülür. Doğum ağırlığı değişkeni için çizilen Schoenfeld artıkları grafiğinde ise, kabul edilebilir bir yatay simetri gözlenmiştir. İki nicel değişkene ilişkin grafikler Şekil 3 ve Şekil 4'deki gibidir.

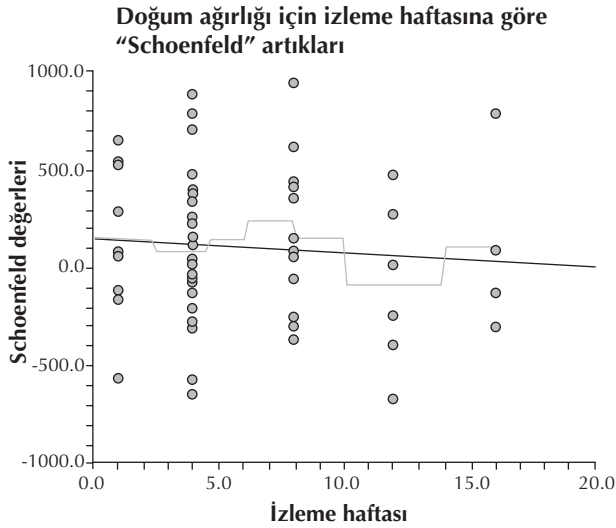
Bilindiği gibi Schoenfeld artığı, değişkenin gerçek değeri ile ağırlıklı risk ortalaması arasındaki fark şeklinde tanımlanmaktadır. Eşantıyon mama hediye edilmesi değişkeni için 0 ve 1 olmak üzere iki kategori söz konusudur. Dolayısıyla bu değişken için artık hesaplarken, gerçek değer yerine geçecek iki ayrı (kategorik) değer olacağından anlamsızlık kaçınılmaz olacaktır. Bu nedenle kategorik değişkenler için Schoenfeld grafiğinin çizilmesi önerilmektedir.



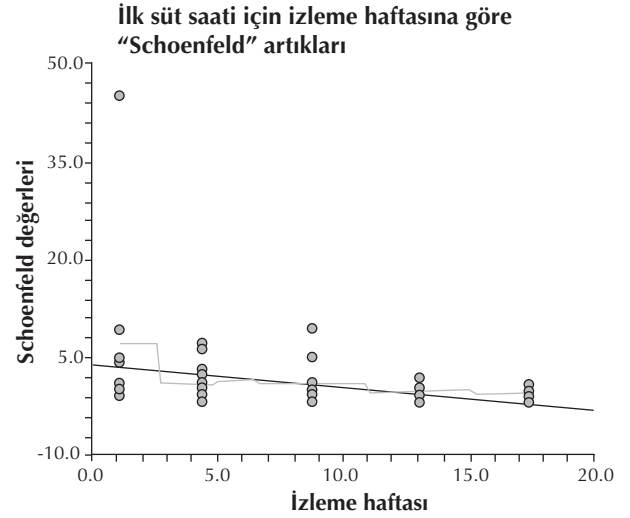
Şekil 1. Bebeğe hastanede mama verilmesi değişkenine ilişkin yaşam fonksiyonu grafiği.



Şekil 2. Bebeğe hastanede mama verilmesi değişkenine ilişkin risk fonksiyonu grafiği.



Şekil 3. Doğum ağırlığı değişkenine ilişkin Schoenfeld artıkları grafiği.



Şekil 4. İlk süt saati değişkenine ilişkin Schoenfeld artıkları grafiği.

TARTIŞMA

Uluslararası çalışmalar başta olmak üzere, Cox regresyon analizi son yıllarda tıp alanında çok yaygın bir şekilde kullanılmaktadır. Bu çalışmada da Cox regresyon analizi genel hatlarıyla irdelenerek, Schoenfeld artıklarına ilişkin analiz yapılmıştır.

İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Çocuk Sağlığı Anabilim Dalı Yeni Doğan Ünitesinde izlenen 100 olgu üzerinde yapılan Cox Regresyon analizi sonuçlarına

göre bebeğin süt emmeyi kesmesini, bebeğin doğum ağırlığı, ilk süt verme zamanı ve bebeğe hastanede mama verilmesi değişkenlerinin etkilediği görülmüştür. Bu sonuca göre, bebeğin doğum ağırlığı arttıkça ve ilk süt verme zamanı geciktikçe, bebeğin süt emmeyi kesmesi olasılığını arttırmaktadır. Bunun yanı sıra, hastanede bebeğe mama verilmesi de anılan olasılığı artırıcı yönde etkilidir.

Teşekkür

"Factors influencing the duration of exclusive breastfee-

ding in a group of Turkish women". J Hum Lact. 2001 adlı makalenin [13] yazarlarına araştırma verilerini kullanmamıza izin verdikleri için teşekkür ederiz.

KAYNAKLAR

1. Cox, D.R., Regression Models and Life-Tables, Journal of the Royal Statistical Society, 1972 Series B, Vol.34, No: 2, 187-220.
2. Hougaard, P., Analysis of Multivariate Survival Data, Springer-Verlag, New York. 2000.
3. Harrell, F.E., Regression Modelig Strategies, Springer-Verlag, New York. 2001.
4. Iachine, I., Basic Survival Analysis, Erişim:[http://www.biostat.sdu.dk/courses/e02/basalebegreber/bb_sur_e01sm.pdf] Erişim Tarihi: 03.05.2006
5. Özdamar, K., SPSS ile Biyoistatistik, Kaan Kitabevi, Eskişehir. 2001.
6. FOX, J., Cox Proportional-Hazards Regression for Survival Data, 2002 Erişim: [<http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-cox-regression.pdf>] Erişim Tarihi: 03.05.2006.
7. Therneau, T.M., Grambsch, P.M., Modeling Survival Data:Extending The Cox Model, Springer-Verlag, New York. 2000.
8. Lee, E.T., Wang, J.W., Statistical Methods for Survival Data Analysis, John Wiley&Sons, New Jersey. 2003.
9. Walters, S.J., What is a Cox model?, Erişim: http://www.jr2.ox.ac.uk/bandolier/painres/download/whatis/COX_MODEL.pdf Erişim Tarihi: 03.05.2006.
10. Terzi, Y., Cengiz, M.A., Bek, Y., Cox Regresyon Modelinde Oransal Hazard Varsayımının Artıklarla İncelenmesi ve Akciğer Kanseri Hastaları Üzerinde Uygulanması, Türkiye Klinikleri, J Med Sci, 2003, 25:770-775.
11. Schoenfeld, D., Partial residuals for the proportional hazards regression model. Biometrika 1982; 69: 239-241.
12. Collett, D., Modelling Survival Data in Medical Research, Chapman&Hall, London. 1994.
13. Alikışıfoğlu M., Erginöz E., Gür ET., Baltaş Z., ve ark. Factors influencing the duration of exclusive breastfeeding in a group of Turkish women. J Hum Lact 2001 Aug;17(3):220-6.
14. Sertkaya, D., Ata, N., Sözer M.T., Yaşam Çözümlemesinde zamana bağlı açıklayıcı değişkenli Cox regresyon modeli, Ankara Üniversitesi Tıp Fakültesi Mecmuası, 2005,58:153-158.
15. Şenocak, M., Biyoistatistik, İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Yayınları. İstanbul, 1998.