Istanbul Commerce University Journal of Science
İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 16(31), Bahar 2017
http://dergipark.gov.tr/ticaretfbd

Review Article

SAYMA SÜREÇLERİYLE YAŞAM SÜRDÜRME ANALİZİ*

Emel BAŞAR[1]

Gazi Üniversitesi, Fen Fakültesi İstatistik Bölümü
ebasar@gazi.edu.tr

Öz

Yaşam sürdürme analizi (survival analysis), iyi tanımlanmış bir başlangıç noktasından itibaren ölçülen ve önceden tanımlanmış herhangi bir olayın gerçekleşmesine kadar geçen zaman sürelerinden oluşan verinin analizinde kullanılmaktadır. Yaşam sürdürme zamanlarından oluşan yaşam sürdürme verisi, çoğu zaman durdurmaya (censoring) bağlı olarak tamamlanamamıştır ve bilinen istatistiksel yöntemler kullanılamamaktadır. Yaşam sürdürme fonksiyonu ve hazard hızı, bütün yaşam sürdürme analizi teorisinde açıklanmaya çalışılan iki temel kavram olmaktadır. Bu kavramları açıklamak üzere özellikleri birbirlerinden farklı pek çok model kurulmaktadır. Bu çalışmada, temel olarak hazard hızını modelleyen ve yaşam sürdürme analizinde yaygın olarak kullanılan bazı modeller sayma süreçleri çerçevesinde incelenmeye çalışılmıştır.

Anahtar kelimeler: Yaşam sürdürme analizi, hazard hızı, Cox model, sayma süreci, toplamsal hazard modeli, çarpımsal-toplamsal hazard modeli, kırılganlık**.**

Tarama Makalesi

COUNTING PROCESS IN SURVIVAL ANALYSIS

Abstract

The survival analysis is used to analyzing survival data which are measured from well defined starting point to the occurrence of well defined any event. Survival data are consisting of the survival times and they are frequently incomplete observations due to censoring. So, ordinary statistical methods cannot be used. The survival function and the hazard rate are two basic concepts to be explained in theory of survival analysis. There are many models with different properties to explain these concepts. In this study, in the context of counting process some models that are used in the survival analysis have been considered and basically have been explained hazard rate.

Keywords: Survival analysis, hazard rate, Cox's model, counting process, additive hazards model, multiplicative-additive hazards model, frailty.

* Received / Geliş tarihi: 25/10/2016 Accepted / Kabul tarihi: 18/01/2017
[1] Corresponding Author/ Sorumlu Yazar : ebasar@gazi.edu.tr

1. GİRİŞ

Yaşam sürdürme analizi istatistiğin en eski alanlarından biridir. Bu konudaki ilk çalışmalar 17. yüzyılda demografi alanında yaşam tablosu kuruluşuyla başlamış, daha sonra demografların ve aktüerlerin katkıları ile geliştirilmiştir. Bu bağlamda günümüzde geliştirilerek kullanılmaya devam eden ilk yaşam tablosu 1662 yılında İngiliz istatistikçi John Graund (1620-1674) tarafından yapılmıştır. Başlangıcından itibaren İkinci dünya savaşının sonuna kadar aktüerler tarafından geliştirilen klasik yaklaşımlar yaşam sürdürme analizine egemen olmuştur (Andersen ve Keiding,1998). Yaşam sürdürme analizi ilk zamanlarda, tam olarak adının da işaret ettiği gibi, ölüm hızları ya da ölümlülük gibi gerçekten yaşamın sürdürülmesine ilişkin olarak düşünülmekteyken, günümüzde yalnızca ölümle ilgili olmayıp, çalışılmak istenen herhangi bir türdeki olayın meydana gelmesi zamanlarının analizi olarak çok geniş bir anlama sahiptir. Mühendislikte güvenilirlik analizi (reliability analysis), ekonomide süre analizi (duration analysis), sosyolojide olay geçmişi analizi (event history analysis) olarak isimlendirilmektedir.

Yaşam sürdürme verisinin önemli bir özelliği diğer veri türlerinde olmayan durdurmanın (censoring) bulunmasıdır. Yani ilgilenilen olayın, gözlemin yapıldığı zaman süresi boyunca mutlaka gözlemlenmiş olması gerekli değildir. İlgilenilen olay bazı birim ya da bireyler için gerçekleşir, ancak bazıları için gerçekleşmez. Kısaca, bir birim ya da bireye ilişkin sürecin gözlemlenmesinin belli bir zamanda durdurulmasıdır ve bu andan sonra ne olduğuna ilişkin daha fazla bilgi yoktur. Bu durumda yaşam sürdürme verisi tamamlanmamıştır. Tamamlanmış ve tamamlanmamış gözlemlerin bir arada bulunması yaşam sürdürme verisinin temel karakteristiği olmaktadır. Bu nedenle durdurulmuş gözlemlerin de bulunduğu verinin analizi için bilinen istatistiksel yöntemler yetersiz kalmakta ve özel yöntemlere ihtiyaç duyulmaktadır.

Yaşam sürdürme alanında 1950'li yıllara kadar klasik yaklaşımlar kullanılmış, asıl gelişme bu tarihten sonra meydana gelmiştir. Kaplan-Meier tarafından önerilen, ünlü yaşam sürdürme eğrisi tahminin yer aldığı çalışma yeni dönemin başlangıcı olmuştur (Kaplan-Meier, 1958). Bu çalışma, "ISI Web of Knowledge"da 33000'den çok alıntıyla istatistik tarihinde en çok alıntı yapılan makalelerden biri olmuştur (Aalen vd., 2009). Klasik yaşam tablosu yöntemi zaman aralıklarını bir, beş gibi sabit ve geniş aralıklara bölerken, Kaplan-Meier yöntemi oldukça küçük aralıklarda çalışmaktadır ve aslında aralıklar sonsuz küçük değerdedir. Böylece bir bakıma eski yaşam tablosunun sürekli-zamana ilişkin olan biçimi önerilmiştir. Önerileri, gün be gün hastaların izlendiği ve ilgilenilen olayın meydana geldiği zamanlarının kaydedildiği klinik çalışmalardan elde edilen, yeni türde yaşam sürdürme verisinin gelişmesine yol açmıştır. Ancak böyle çalışmalardaki birey sayısı demografik ve aktüeryal çalışmalardaki birey sayısından çok daha az olmaktadır. Böylece Kaplan-Meier yönteminin gelişmesiyle yeni tür verinin yarattığı yeni problemlere yanıt bulunabilmektedir.

Kaplan-Meier'in makalesi 1958 yılında yeni bir alan açmış, ancak pek çok soruyu da beraberinde getirmiştir. Örneğin yaşam sürdürme eğrileri karşılaştırılabilir mi? gibi sorular, yeni çalışmaların yapılması yolunu açmıştır. İki ya da daha çok yaşam sürdürme eğrisinin karşılaştırılmasına ilişkin testler, 1960'lı ve 1970'li yıllar boyunca geliştirilmiştir. Zaman kavramıyla ölçülen değişkenin üzerinde etkili olabilecek eşdeğişkenlerin (covariates) modele nasıl ekleneceği sorusu ilk defa 1972'de David Cox tarafından önerilen orantılı hazard modeli (proportional hazards model) ile çözülmüştür (Cox, 1972). Cox'ın makalesi önemli bir gelişmeye neden olmuştur ve ispatı, "ISI Web of Knowledge"da 24000'den çok alıntıyla çok büyük bir etkiye yol açmıştır (Aalen vd., 2009).

Kaplan-Meier tahmin edicisinin asimptotik özelliklerinin neler olduğu ya da Cox modelinin uygun olup olmadığı gibi soruların ortaya çıkması başka teori arayışlarına neden olmuştur. Yaşam sürdürme verisinin yapısı gereği stokastik süreç özelliği göstermesi, martingale kavramının, genel bir teori arayışı içinde yer almasını sağlamıştır. Martingale yöntemleri, karmaşık istatistiksel kavramlar için basit ifadeler elde etmeyi, test istatistikleri ve tahmin ediciler için asimptotik dağılımların elde edilmesini ve durdurulmuş veriye ilişkin işlevsel özellikleri incelemeyi mümkün kılmaktadır. Bu gelişmelerin başlangıcı 1980'li yılların başına dayanmaktadır. Bu alanda ilk çalışma Aalen tarafından yapılmıştır (Aalen, 1978b; 1980). Andersen ve Gill; Fleming ve Harrington; Andersen, Borgan, Gill ve Keiding; Aalen, Borgan ve Gjessing'in çalışmalarının, teorinin gelişmesine önemli katkısı olmuştur (Aalen vd., 2009; Andersen ve Gill 1982; Andersen vd., 1993; Fleming ve Harrington, 2005).

Aslında Martingale kavramı resmen olmasa bile yaşam sürdürme alanında gizli olarak yer almıştır. 1959'da önerilen Mantel-Haenszel testinde, 1975'de Cox'ın önerdiği kısmi olabilirlik fonksiyonunun elde edildiği çalışmada, 1977'de Tarone ve Ware'in önerdiği testte martingale düşüncesi sezgisel olarak yer almıştır (Cox, 1975; Mantel-Haenszel, 1959; Tarone ve Ware, 1977).

2. YAŞAM SÜRDÜRME FONKSİYONU VE HAZARD HIZI

Yaşam sürdürme analizi teorisinde önemli olan iki temel kavram bulunmaktadır. Bunlar yaşam sürdürme fonksiyonu ve hazard hızıdır. Yaşam sürdürme eğrisi belli bir zamana kadar kaç kişinin hayatta kaldığı bilgisini verirken, hazard hızı, zamanın bir fonksiyonu olarak önceden meydana gelmemesi koşulu altında incelenen olayın gerçekleşmesi riskini vermektedir.

Bir birim ya da bireye ilişkin yaşam sürdürme sürelerinden oluşan rastgele değişken T'lerin aynı dağılımlı ve bağımsız (iid), pozitif değerler alan bir rastgele değişken olduğu varsayılmaktadır. Yaşam sürdürme fonksiyonu S(t), t zamanına kadar ilgilenilen olay henüz gerçekleşmemiş bireyler için, olayın gerçekleşmesine ilişkin beklenen oranı vermektedir. İlgilenilen olayın t zamanına kadar gerçekleşmemesinin koşulsuz olasılığıdır ve aşağıdaki gibi gösterilmektedir.

$$S(t) = P(T > t) \tag{1}$$

Hazard hızı $\alpha(t)$, koşullu olasılık olarak tanımlanmaktadır. Rastgele değişken T'lerin sürekli olduğu varsayılarak, t zamanına kadar ilgilenilen olayın henüz gerçekleşmediği gözlemler dikkate alındığında, $(t, t+dt)$ gibi küçük bir zaman aralığında ilgilenilen olayın meydana gelmesi olasılığı $\alpha(t)dt$ olmaktadır. Böylece hazard hızı limit olarak aşağıdaki gibi tanımlanmaktadır.

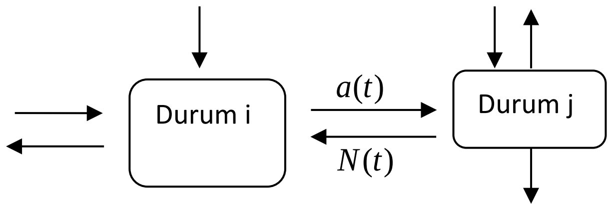
$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} p(t \le T < t + dt \mid T \ge t)$$

Yaşam sürdürme eğrisi birden başlayan, sonra zaman boyunca artmayan bir fonksiyon iken, hazard fonksiyonu negatif olmayan herhangi bir fonksiyon olabilmektedir. Hazard hızı basit gibi görünmekle birlikte aslında anlaşılması zor bir kavramdır. Durdurulmuş yaşam sürdürme verisinden K-M tahmin edicisiyle yaşam sürdürme eğrisi kolayca tahmin edilirken, zamana bağlı keyfi bir fonksiyon olan hazard hızını tahmin etmek daha güç olmaktadır. Çoğu zaman aşağıda tanımlanan birikimli hazard hızını tahmin etmek daha kolay olmaktadır.

$$A(t) = \int_0^t \alpha(s)ds$$

Bu eşitlik Nelson-Aalen tahmin edicisi adını almaktadır ve $A(t)$ 'nin parametrik olmayan bir tahmini Nelson tarafından 1972 yılında önerilmiş, Aalen tarafından geliştirilmiştir (Aelen, 1978b; Nelson, 1972).

Yaşam sürdürme analizinde martingale kavramı, Nelson-Aalen tahmin edicisiyle birlikte kullanılmaya başlanmıştır. Tahmin süreci hazard hızı kavramına dayanmaktadır. İlgilenilen olayların meydana gelmesi iki durumlu bir Markov zinciri gibi düşünülmektedir. Durumlar ve geçişler Şekil 1'de yer almaktadır.



Şekil 1. Markov Zincirinde Geçişler (Aalen vd., 2009)

t zamanında i'nci durumda olan bireylerin sayısı $Y(t)$, $[0, t]$ zaman aralığında i'nci durumdan j'inci duruma geçenlerin sayısı $N(t)$ ile gösterilsin. Durum uzayında i ve j durumlarını dikkate alalım. Yeni bir olayın gerçekleşmesine ilişkin hız, yani yeni bir geçiş olması, t zamanındaki bireylerin sayısının, $\alpha(t)$ ifadesi ile çarpımına eşit olacaktır. Böylece geçiş olasılıkları hazard hızı kavramına dayalı olarak elde edilmektedir. Ayrıca yaşam sürdürme verisine ilişkin bir özellik olan durdurma, kolayca bu düzene dâhil edilebilmektedir.

Aalen'in bu düşüncesi "çarpımsal yoğunluk modeli (multiplicative intensity model)" adı verilen genel bir modellemenin başlangıcı olmuştur. Başka bir deyişle sayma süreci $N(t)$'nin hızı olarak tanımlanan $\lambda(t)$, gözlenen bir $Y(t)$ süreci ile bilinmeyen hız fonksiyonu $\alpha(t)$ 'nin çarpımı olarak,

$$\lambda(t) = a(t)Y(t) \tag{4}$$

şeklinde yazılmaktadır. Yaklaşık olarak;

$$dN(t) \approx \lambda(t)dt = a(t)Y(t)dt \tag{5}$$

$$\frac{dN(t)}{Y(t)} \approx a(t)dt \tag{6}$$

olur ve $A(t) = \int_0^t \alpha(s)ds$ 'nin uygun bir tahmini;

$$\int_0^t \frac{dN(s}{Y(s)} \tag{7}$$

olarak elde edilir. Bu ifade tam olarak Nelson-Aalen tahminidir. Yukarıda anlatılan yapı Markov zinciri yapısından çok bir nokta süreci (point process) ya da sayma süreci (counting process) özelliğine sahiptir ve matematiksel temeli Aalen tarafından ortaya konulmuştur (Aelen, 1978b).

Martingale teorisine dayalı olarak bir sayma süreci iki kısım halinde yazılabilmektedir.

$$N(t) = \Lambda(t) + \mathrm{M}(t) \tag{8}$$

$$(Gözlem = Sinyal + Gürült\text{ü})$$

$\Lambda(t)$ birikimli yoğunluk sürecidir ve $\Lambda(t) = \int_0^s \lambda(s)ds$ olarak tanımlanmaktadır. Böylece sayma sürecinden yoğunluk sürecinin integrali çıkarılırsa bir martingale elde edilmektedir. Bu da sezgisel olarak ulaşılan yoğunluk süreci kavramına eşit olmaktadır ve Nelson-Aalen tahmini aşağıdaki gibi yazılabilmektedir.

$$dN(t) = a(t)T(t)dt + dM(t) \tag{9}$$

$$\frac{1}{Y(t)}dN(t) = a(t) + \frac{1}{Y(t)}dM(t) \tag{10}$$

$$\int_0^t \frac{dN(s)}{Y(s)} = \int_0^t a(s)ds + \int_0^t \frac{dM(s)}{Y(s)} \tag{11}$$

Böylece sıfır ortalamalı bir martingale ile $\hat{A}(t)$ yansız bir tahmin edici olmaktadır.

Nelson-Aalen tahmin edicisi $a(t)$ hazard fonksiyonunu esas almaktadır. Burada $\alpha(t)dt$, t zamanında riskte olan bir birey için, bundan sonraki çok küçük $[t, t+dt)$ aralığında olayın meydana gelmesi olasılığıdır.

Bazen hazard fonksiyonu yerine $(s, t\ ]$ sonlu aralığında bir olayın meydana gelmesi olasılığı olan yaşam sürdürme fonksiyonu ile ilgilenilebilir. Nelson-Aalen tahmin edicisini yaşam sürdürme fonksiyonu $S(t)$'nin bir tahminine dönüştürmek için product-integral dönüşümü kullanılır (Gill ve Johansen,1990; Gill, 2005).

$$S(t) = \coprod_{(0,t)} \{1 - dA(s)\} \tag{12}$$

$A(t) = \int_0^t a(s)ds$ hazard fonksiyonu $a(t)$ için birikimli yoğunluktur ve

$$\coprod_{(0,t)} \{1 - dA(s)\} = \exp(-\int_0^t a(s)ds) \tag{13}$$

şeklinde ifade edilebilir. Eğer birikimli yoğunluk kesikli bir ölçüm ise;

$$A(t) \sum_{s_i \le t} h_j \quad (s_1 < s_2 < \cdots)$$

olmak üzere, s_j zamanında h_j yüksekliğinde sıçrama yapan bir fonksiyon olacaktır. O zaman;

$$\prod_{(o,1)} \{1 - dA(s)\} = \prod_{s_i \le t} \{1 - h_j\} \tag{14}$$

olur ve tahmin edici;

$$\hat{S}(t) = \coprod_{(0,t)} \left\{1 - d\hat{A}(s)\right\} \tag{15}$$

biçiminde elde edilir. Elde edilen bu ifade 1958 yılında önerilen Kaplan-Meier tahmin edicisidir. Olayın meydana geldiği zamanlarda, $t_j \le t$ için $1 - 1/Y\,(t_j)$ faktörlerinin çarpımına eşittir. K-M tahmin edicisinin martingale gösterimi;

$$\frac{\hat{S}}{S(t)} - 1 = -\int_0^t \frac{\hat{S}(s^-)}{S(s)Y(s)} dM(s) \tag{16}$$

olur. Sağ taraftaki integral sıfır ortalamalı bir martingale ile tahmin edilebilir bir süreçtir. Bu gösterimle K-M tahmin edicisinin özellikleri kolayca ispatlanabilmektedir (Gill, 1980).

Pek çok durdurma türü olmakla birlikte durdurma türleri genel olarak sağdan durdurma adı altında toplanmaktadır. Sağdan durdurma yoğunluk sürecinin biçimini değiştirmemektedir. Ayrıca sayma süreci yaklaşımının önemli bir özelliği de bağımsızlık gerektirmemesidir.

3. REGRESYON MODELLERİ

Yaşam sürdürme analizinde yapılan pek çok çalışmanın amacı, yaşam süresi üzerinde bir ya da daha çok eşdeğişkenin etkisini incelemektir. Durdurulmuş gözlemlerin bulunduğu yaşam sürdürme verisi için çok sayıda farklı biçimleri ve özellikleri olan regresyon modelleri önerilmiştir.

t zamanında $i = 1,2, \dots, n$ olmak üzere n tane birey olduğunu ve i'nci birey için kimi bileşenleri sabit (fixed), kimi bileşenleri zaman bakımından değişken (time-varying) olan eşdeğişken vektörünün $x_i(t) = \left(x_{i1}(t), x_{i2}(t), \dots x_{ip}(t)\right)^T$ olduğu varsayılsın. N_i'ye ilişkin yoğunluk süreci aşağıdaki gibi tanımlanabilir.

$$\lambda_i(t) = Y_i(t)a(t|x_i) \tag{17}$$

$Y_i(t)$, eğer i'nci birey, ilgilenilen olay bakımından t zamanından hemen önce riskte ise 1 değerini, aksi halde 0 değerini alan bir gösterge değişkenidir. i'nci bireye ilişkin yoğunluk ya da hazard hızı $\alpha(t|x_i)$, sabit ya da zaman bakımından değişken olan eşdeğişkenlerin varlığı altında tanımlanır. Bir regresyon modeli elde etmek için $a(t|x_i)$'nin $x_i(t)$'ye ne şekilde bağlı olduğunu tanımlamak gerekir. Bu ilişkiyi parametrik, yarı-parametrik ve parametrik olmayan modeller kurarak belirlemek mümkündür.

3.1. Göreli Risk Regresyon Modelleri

Bu grupta yer alan regresyon modelleri yarı-parametrik modellerdir ve i'nci bireye ilişkin eşdeğişken vektör $x_i(t)$'nin, hazard hızı $a(t|x_i)$ ile aşağıdaki gibi ilişki içinde olduğu varsayılmaktadır.

$$a(t|x_i) = a_o(t)\tau(\beta, x_i(t)) \tag{18}$$

Burada $\tau(\beta, x_i(t))$ göreli risk fonksiyonu, β eşdeğişkenin etkisini gösteren regresyon katsayı vektörü ve $\alpha_0(t)$ tanımlanmamış temel hazard hızıdır. Yukarıda tanımlanan model parametrik olmayan kısım (temel hazard) ve parametrik kısım (göreli risk fonksiyonu) olmak üzere iki parça halindedir ve bu nedenle yarı-parametrik model olarak isimlendirilmektedir. Model, $N_i(t)$ yoğunluk sürecine dayalı olarak aşağıdaki gibi yazılabilmektedir;

$$\lambda_i(t) = Y_i(t)a_0(t)\tau(\beta, x_i(t)) \tag{19}$$

$\tau(\beta,0)=1$ olduğu varsayıldığı zaman, tüm eşdeğişken değerleri sıfıra eşit olan bir bireye ilişkin hazard hızı $a_0(t)$ olmaktadır.

Farklı biçimlerde risk fonksiyonları tanımlanabilmektedir. Risk fonksiyonu $\tau(\beta, x_i(t)) = \exp\left\{\beta^T x_i(t)\right\}$ biçiminde tanımlanırsa, üstel göreli risk modeli elde edilmektedir. Bu model çok bilinen ve yaygın olarak kullanılan Cox regresyon modeli olarak isimlendirilmektedir.

Bağlantı fonksiyonu olarak $\tau(\beta, x_i(t)) = 1 + \beta^T x_i(t)$ biçiminde bir fonksiyon tanımlanırsa model, doğrusal göreli risk modeli adını almaktadır. Bu model, "toplamsal göreli hazard modeli (additive relative hazard model)" olarak da isimlendirilmektedir. Bir başka model, parametrik kısım için dikkate alınan fonksiyonun $\tau(\beta, x_i(t)) = 1 + \exp(\beta^T x_i(t))$ biçiminde olduğu modeldir. Bu modellerde eşdeğişkenler temel hazard hızını toplamalı olarak artırmakta ya da azaltmaktadır. Ancak eşdeğişkenlerin etkisi tanımlanmamış temel hazard hızı ile birlikte ortaya çıkmakta ve eşdeğişkenlerin ayrı ayrı yorumlanması mümkün olmamaktadır. Ayrıca modelin toplamsal bir yapıya sahip olması nedeniyle hazard fonksiyonun pozitif değer alması her zaman sağlanamamaktadır.

Göreli risk regresyon modellerinde iki bireye ilişkin hazard hızları oranlanırsa;

$$\frac{\alpha(t \mid x_2)}{\alpha(t \mid x_1)} = \frac{\tau(\beta, x_2(t))}{\tau(\beta, x_1(t))} \tag{20}$$

ifadesi elde edilir. Eğer tüm eşdeğişkenler zaman bakımından sabit ise bu oran zaman boyunca aynı kalır ve bu özelliğe sahip olan modeller orantılı hazard modelleri olarak isimlendirilir. Özel olarak Cox'ın modeli dikkate alındığı zaman, diğer eşdeğişkenler sabit iken j'inci eşdeğişkendeki bir birimlik artış, hazard hızını e^{β_j} kadar artırmaktadır. Genel anlamda bu değer j'inci eşdeğişkene ilişkin göreli risk olmaktadır. Regresyon katsayıları doğrusal göreli risk modelinde bu kadar kolayca yorumlanamamaktadır. Bu nedenle Cox'ın modeli çok kullanılan bir model olmaktadır (Aalen vd., 2008).

Cox'ın modeli çok esnek bir model olmasına rağmen, eşdeğişken etkilerinin zaman bakımından değişmez olduğunun varsayılması, modelin uygunluğunun denetlenmesini önemli bir problem haline getirmiştir. Modelin uygunluğu konusunda gerek grafiksel olarak, gerek artıklara dayalı pek çok çalışma yapılmıştır.

Orantılılık varsayımının bozulduğu ortaya çıkarsa, Cox'ın modeline zamana bağlı bazı eşdeğişkenlerin eklenmesi mümkündür. Birçok farklı biçimde zamana bağlı eşdeğişken modele dâhil edilebilmektedir. İzleme zamanı farklı zaman aralıklarına bölünerek her bir aralık içinde geçerli olmak üzere parçalı sabit göreli risk varsayımı yapılabilir. Bir başka yol ise, zaman bakımından değişken (time-varying) olan eşdeğişkenin etkisi için parametrik bir fonksiyon tanımlamaktır. Ancak hangi fonksiyonun uygun olduğu konusu çok açık değildir.

Bir diğer yol ise tabakalı modelin kullanılmasıdır. Orantılılık özelliği göstermeyen eşdeğişkenin düzeyleri tabaka olarak kabul edilir. Her tabaka, farklı bir temel hazard fonksiyonuna sahip olmakta ancak regresyon katsayıları bütün yığın için aynı olmaktadır. Modelin dezavantajı, tabakalı olarak alınan eşdeğişkenin etkisinin özetlenmesinin zorluğudur. Çok sayıda tabakanın varlığı durumunda model uyumu zorlaşır ve tahminlerin kestirimleri zayıflar. Ayrıca aynı anda pek çok eşdeğişkene nasıl anlam verileceği güçleşir.

3.2. Toplamsal Regresyon Modelleri

Göreli risk regresyon modellerinde eşdeğişkenlerin etkisinin, temel hazard fonksiyonu üzerinde çarpımsal bir rol oynadığı ve etkilerinin zaman boyutunda sabit olduğu varsayılmaktadır. Ancak bazı durumlarda eşdeğişkenlerin etkisi çarpımsal olmaktan çok toplamsal olmaktadır. Yani başka bir deyişle eşdeğişkenin etkisinin ölçümü, göreli bir ölçüden çok mutlak bir ölçü olarak ortaya çıkmaktadır. Bu modellerde eşdeğişkenlerin etkilerinin zaman boyunca değişmesine izin verilmektedir ve modeller parametrik olmayan modellerdir. Çeşitli biçimlerde kurulan modeller pek çok yazar tarafından incelenmiştir. Yukarıda değinilen toplamsal göreli hazard modelleri kimi zaman bu grup içinde yer almaktadır. Ancak yine de göreli olma özelliği öne çıkmaktadır. Tümüyle parametrik olmayan özelliğe sahip olan model Aalen tarafından önerilen ve kendi adıyla anılan Aalen'in toplamsal hazard modelidir (Aalen's additive hazards model) ve aşağıdaki gibi tanımlanmaktadır (Aalen, 1980).

$$\alpha(t|x_i) = \beta_0(t) + \beta_1(t)x_{i1}(t) + \cdots + \beta_p(t)x_{ip}(t) \tag{21}$$

$$\lambda i(t) = Yi\ (t)(\beta 0(t) + \beta_1(t)x_{i1}(t) + \cdots + \beta_p(t)x_{ip}(t)) \tag{22}$$

Burada $x_i(t) = (x_{i1}(t), x_{i2}(t), \ldots x_{ip}(t))^T$ t zamanında i'nci bireye ilişkin eşdeğişken vektörü, $\beta_0(t)$ bütün eşdeğişken değerleri sıfır olan bir bireye ilişkin temel hazard hızını göstermektedir. Bu modeldeki katsayılar, t zamanında, eşdeğişkendeki bir birimlik değişmenin temel hazard hızı $\beta_0(t)$'de meydana getirdiği değişmeyi göstermektedir. Bu anlamda $\beta_j(t)$ t zamanında j'inci eşdeğişkenin etkisini gösteren risk artışıdır (excess risk). Diğer modellerden farkı, Aalen'in modelinde, zaman boyutunda $\beta_j(t)$ katsayılarının büyüklüğünde ve işaretinde serbestçe değişiklik olmasına olanak verilmesidir. Ayrıca eşdeğişkenlerin etkileri de zaman boyunca değişebilmektedir. Temel hazard hızı tanımlanmamış bir fonksiyondur ve model tam olarak parametrik olmayan bir modeldir.

En çok olabilirliğe dayalı tahmin tekniklerinin kullanıldığı orantılı hazard modellerinin aksine, risk katsayılarının tahminleri en küçük kareler tekniklerine dayalı olarak elde edilmektedir. Aalen, modeldeki $\beta_j(t)$ katsayılarını doğrudan tahmin etmek yerine birikimli regresyon katsayılarının en küçük kareler tahminini elde etmiştir.

Yarı parametrik toplamsal regresyon modelleri Lin ve Ying; McKeague ve Sasieni tarafından önerilmiştir (Lin ve Ying, 1994; McKeague ve Sasieni, 1994). Lin ve Ying tarafından önerilen model;

$$\lambda_i(t) = Y_i(t)\left\{\beta_0(t) + \gamma x_i(t)\right\} \quad (23)$$

biçimindedir. Burada $\beta_0(t)$, tanımlanmamış temel hazard hızı, γ, zaman boyutunda sabit olduğu varsayılan regresyon katsayı vektörü, $x_i(t)$ eşdeğişken vektörüdür.

McKeague ve Sasieni tarafından önerilen model ise, bir bakıma Aalen'ın modeli ile Lin ve Ying tarafından önerilen modelin karışımı olarak düşünülebilir.

$$\lambda_i(t) = Y_i(t)\left\{\beta(t) z_i(t) + \gamma x_i(t)\right\} \quad (24)$$

Bu modelde regresyon katsayılarının bazıları zaman bakımından değişebilirken bazılarının sabit olduğu varsayılmaktadır.

Toplamsal ve çarpımsal modelleri birlikte ele alan ve çarpımsal-toplamsal modeller (multiplicative-additive models) olarak adlandırılan farklı modeller de kurulabilmektedir.

3.3. Parametrik Regresyon Modelleri

Parametrik olmayan modeller daha çok biyoistatistik ve tıp alanında kullanılırken, parametrik modeller mekanik bir parçanın ömür süresiyle ilgili güvenirlik çalışmalarında kullanılmaktadır. En çok kullanılan dağılımlar üstel, Weibull dağılımları olmaktadır.

Daha önce $[o, t)$ aralığında, i'nci bireye ilişkin tanımlanan olayın meydana gelmesi sayısı bir sayma süreci olan $N_i(t)$ ile gösterilmişti. Olay, her bir birey için en fazla bir kez ortaya çıkmakta ve $N_i(t)$ yalnızca 0 ya da 1 değerini almaktadır. Tekrarlı olay (recurrent event) verisinde ise her birey için olay birden fazla kez tekrarlanabilir ve $N_i(t)$ 1'den büyük tam sayılı değerler alabilir. Bu durumda

$N_i(t)$ 'nin yoğunluk süreci q-boyutlu parametre ile $\lambda_i(t;\theta)$ olarak gösterilmekte ve aşağıdaki gibi tanımlanmaktadır;

$$\lambda_i(t;\theta) = Y_i(t)\alpha_i(t;\theta) \tag{25}$$

Burada, $Y_i(t)$ i'nci birey için; t zamanından hemen önce birey gözlem altında ise 1, aksi halde 0 değerini alan bir gösterge değişkeni, $\alpha_i(t;\theta)$, yoğunluk ya da hazard hızı olmaktadır.

Regresyon modeli için hazard hızı $\alpha_i(t;\theta)$, bileşenleri sabitlenmiş ya da zaman bakımından değişken olan eşdeğişken vektörü $x_i(t) = (x_{i1}(t), x_{i2}, ..., x_{ip}(t))^T$ 'ye bağlı olacaktır.

Bu modeller, gözlenebilen eşdeğişkenlere bağlı olarak ortaya çıkan ve bireyler arasındaki farklılıkları modellemek için kullanılabilir. Ancak, genetik ya da çevresel faktörlere bağlı olarak ortaya çıkan, gözlenemeyen değişkenlere bağlı olan bireyler arasındaki farklılıklar modellenmek istenirse başka bir modelleme yapmak gerekmektedir. Bu problem, kırılganlık (frailty) modelleri kurularak çözülebilmektedir.

4. KIRILGANLIK MODELLERİ

Bireyler birbirlerinden farklıdır. Bu durum yaşamın ve istatistiğin temel bir gözlemi olmaktadır. Bazen bir birey için yararlı olan bir işlemin, başka bir birey için daha az yararlı olacağı sonucuna ulaşılabilir. Bir bireye ilişkin yüksek bir risk faktörü, başka bir birey için daha düşük bir risk anlamına gelebilir. Bu değişim her yerde gözlenebilen bir değişimdir.

İstatistiksel analizin amaçlarından biri de böyle bir değişimi belirleyen faktörleri doğru biçimde anlamaktır. Bu nedenle yapılan bir regresyon analizinde, gözlenebilen eşdeğişkenlerle açıklanamayan bazı değişimler bulunmaktadır ve açıklanamayan bu değişim önemli olabilmektedir.

Yaşam sürdürme analizinde, şimdiye kadar anlatılan modeller, gözlenebilen eşdeğişkenlere bağlı olarak bireyler arasında görülen farklılıkları modellemek için kullanılabilir. Eşdeğişkenlerle açıklanamayan, değişimi gözlenemeyen heterojenlik (unobserved heterogenity), modellenmek istendiği zaman kırılganlık modelleri söz konusu olmaktadır. Yani bazı bireyler için tanımlanan olayın gerçekleşmesi daha muhtemeldir ve bireyler arasında temel bir rastgelelik (radomness) daima mevcut bulunmaktadır.

Kırılganlık sözcüğü, gözlenemeyen heterojenliği tanımlamak üzere ilk olarak Vaupel ve arkadaşları tarafından tek değişkenli yaşam sürdürme modellerinde kullanılmıştır . Vaupel, demografik anlamda bir bireye ilişkin ölüm gücünü (force of mortality); bireyin zayıflık düzeyi ve temel bir fonksiyonun çarpımı olarak tanımlayan bir rasgele etki modeli önermiştir (Vaupel vd., 1979). Clayton ve Cuzick temel hazard fonksiyonu üzerinde çarpımsal biçimde etkili olan orantılı hazard kırılganlık modelini önermiştir (Clayton ve Cuzick, 1985). Hougaard bireyler arasındaki heterojenliği gamma, ters Gaussian dağılımlarını kullanarak parametrik olarak modellemiştir (Hougaard 1984, 1986). Sinha ve Gustafson çok düzeyli çok değişkenli yaşam sürdürme verisi için bayesci hiyerarşik zayıflık modellerini incelemiştir (Gustafson, 1995; Sinha, 1993). Yau rastgele etkiler için çok düzeyli (multilevel) modelleri önermiştir (Yau, 2001). Yashin ve arkadaşları ilişkili veri yapısında kırılganlık modelini incelemiş (Yashin vd., 1995), Huber ve Vonta modeli, veride durdurma ve budanma (truncation) durumu varken ele almıştır (Huber ve Vonta, 2004).

Çözümlenmek istenen problemlerde kırılganlığın var olması, pek çok birbirinin içine girmiş hataya yol açmakta ve yanlış sonuca ulaşmaya neden olabilmektedir. Söz konusu olay, yüksek riske sahip olan bireylerde, diğer bireylerden daha önce meydana gelmektedir. Böylece bir süre sonra risk setinde kalanlar daha düşük ortalama kırılganlığa sahip olacaklardır. Bu durum aşağı doğru hareket eden bir hazard hızı etkisi izlenimi vermektedir. Yığına ilişkin çok yavaş artan hatta azalan bir hazard hızı, bireysel riskin de azaldığı biçiminde yorumlanabilmektedir. Ancak kimi zaman bu durum bütünüyle yanlış olabilmektedir. Yığın için gözlemlenen hazard hızı azalırken her bir bireye ilişkin gözlemlenemeyen hazard hızının artması mümkün olabilmektedir.

Bu durum aslında hazard hızını gerçek anlamının ne olduğu sorusunu ortaya çıkarmaktadır ve hazard hızı, modellerin kurulmasında, hatta daha geniş bir anlamda sayma süreci yaklaşımının kuruluşunda önemli bir kavaram olmaktadır. Hazard hızının zaman boyutunda gelişmesi, bireye ilişkin riskteki gelişmeyle yarışan riskler tanımlanmaktadır. Böylece hazard hızının biçimi gözlendiği zaman, bunun bireylerin içindeki değişime mi yoksa bireyler arasındaki farklılığa mı bağlı olduğunu bilmek zor olmaktadır.

Daha önce de değinildiği gibi, bazen söz konusu olan olay bir birey için birden fazla sayıda gerçekleşebilir. Böyle olaylar tekrarlı olaylar adını almaktadır. Başka bir durumda ise ilgilenilen birden çok olay bulunmaktadır. Böyle olaylar yarışan riskler (competing risks) olarak adlandırılmaktadır. Her birey için olayın birden çok kere gerçekleşebilmesi durumu çok-durumlu modeller ile tanımlanmaktadır. Tekrarlı olaylar ya da başka tür çok değişkenli yaşam sürdürme verisi varken rastgele etki modellerinin kuruluşunda kırılganlık kavramından yararlanılmaktadır. Böylece kırılganlık teorisi çeşitli biçimlerdeki rastgele etki modellerinin kurulmasında önemli olmaktadır.

Uygulamada kırılganlık kavramının ne anlama geldiği yani heterojenlik kaynağının ne olduğu birkaç maddede özetlenebilir. İlk olarak başlangıçta var olan biyolojik farklılıklardan bahsedilebilir. Örneğin bazı bireyler hassas bir kalple ya da kansere genetik bir yatkınlıkla doğar. İkinci olarak kırılganlık, dinamik bir kavaram olarak, örneğin stresli bir yaşamın zaman boyunca bireyleri değiştirebilmesi olarak ortaya çıkabilir. Üçüncüsü de geç kalınmış ya da erken teşhis edilmiş bir hastalığa bağlı olarak ortaya çıkan heterojenliği kapsayabilir. Bu durum biyolojik özelliklerin doğasında bulunan kırılganlıktan biraz daha farklı olmaktadır. Bu anlamda kırılganlık, bireyler arasındaki pek çok farklı heterojenlik türünü kapsaması bakımından karmaşık bir kavram olmaktadır. Kısaca bu kavram, bireyler arasında farklılıklar olduğunu varsaymaktadır. Ancak bu farklılıkların risk bakımından nasıl değerlendirilmesi gerektiği önemli bir soru olarak karşımıza çıkmaktadır.

Kırılganlık kavramını matematiksel olarak ifade etmenin yollarından biri, her bireye ilişkin hazard hızı üzerinde orantılı biçimde hareket eden bir kırılganlık faktörü Z'ye sahip olduğunun varsayılmasıdır. Böylece kırılganlık faktörü Z'ye sahip olan bir bireye ilişkin hazard hızı;

$$a(t|Z) = Z\alpha(t) \tag{26}$$

olmaktadır. Burada, kırılganlık düzeyini tanımlayan Z, bireylerin ait olduğu yığına ilişkin bir rastgele değişken olarak kabul edilmektedir. Z ve $\alpha_0(t)$ ifadelerinin her ikisi de gözlenemez (unobservable). Bir yığından gözlenebilen şey bireye ilişkin hazard hızı değildir, ancak farklı kırılganlıklara sahip olan çok sayıda bireye ilişkin bir sonuçtur veya yığın hazard hızı olarak adlandırılan kavramdır.

Z verildiğinde t zamanına kadar yaşam sürdürme olasılığı, $A(t) = \int_0^t \alpha(u)du$ olmak üzere;

$$S(t|Z) = e^{-ZA(t)} \tag{27}$$

olarak tanımlanmaktadır. Yığına ilişkin yaşam sürdürme fonksiyonu ise;

$$S(t) = E\left\{e^{-ZA(t)}\right\} \tag{28}$$

olmaktadır. Z'ye ilişkin Laplace dönüşümü $\mathcal{L}(c) = E(\exp(-cZ))$ ile tanımlanırsa yaşam sürdürme fonksiyonu;

$$S(t) = L(A(t)) \tag{29}$$

ve $\mu(t)$ ile gösterilen yığın hazard hızı;

$$\mu(t) = \alpha(t)\frac{-L'(A(t))}{L(A(t))} \tag{30}$$

biçiminde elde edilmektedir. Böylece yığın ve bireye ilişkin hazard hızı arasındaki fark ikinci terimle belirlenerek daima azalan bir fonksiyon olduğu gösterilebilmektedir. Model, bireye ilişkin heterojenliği açıklamanın zamana bağlı olmayan basit bir ifadesi olmakla birlikte heterojenlikle ilgili sorunun çözümünde etkili bir model olmaktadır. Laplace dönüşümünün sağlandığı dağılımlar kırılganlık dağılımları olarak alınabilmektedir. Bu konuda gamma dağılımı genel bir seçim olarak kabul edilmektedir. Gamma dağılımından başka pozitif kararlı (positive stable), güç varyans fonksiyonu (power variance function), lognormal, ters (inverse) Gaussian, Weibull ve bileşik (compound) Poisson dağılımlarından yararlanılmaktadır (Hanagal, 2011; Wienke, 2011).

5. SONUÇ

Günümüzde hemen her alanda istatistiksel analizlerden yararlanılmaktadır. Analizlerin kullanımı öncelikle elde edilen verinin yapısına ve verinin doğasında bulunan değişkenliğe bağlıdır. Rastgelelik, veriye ilişkin değişimin hepsini kapsayan yakalanması zor, geniş bir kavramdır. İstatistiksel analizlerin başlıca amacı, rastgeleliği yaratan faktörleri doğru biçimde anlamak ve yorumlamaktır. Bu anlamda rastgelelik her durumda incelenmesi gereken bir olgu olmaktadır.

Bu çalışmada yaşam sürdürme analizindeki gelişmeler tarihsel boyut dikkate alınarak incelenmiş, yaşam sürdürme analizinin temel kavramları, ortaya çıkan yeni problemler ve problemleri çözmek üzere ileri sürülen yeni düşünceler sayma süreçleri kuramı bakımından açıklanmaya çalışılmıştır. Yaşam sürdürme analizinde yer alan modeller, genel özellikleri dikkate alınarak sınıflanmaya çalışılmış ve tüm modeller sayma süreçleri kuramı bağlamında ele alınarak incelenmiştir.

KAYNAKLAR

Aalen, O. O., (1978b), “Nonparametric inference for a family of counting processes”, Annals of Statistics, 6, 701–726.

Aalen, O. O., (1980), “A Model for Nonparametric Regression Analysis of Counting Process. In Lectures notes on Mathematical Statistics and Probability”, Klonecki W., Kozek A., Rosinski J., eds. Springer-Verlag, New York, 1-25.

Aalen, O. O., Borgan, Ø., Gjessing, H., K., (2008), Survival and Event History Analysis: A Process Point of View. Springer-Verlag, NewYork.

Aalen, O. O., Anderson, P., Borgan, Ø., Gill, R., Keiding, N., (2009), “History of applications of martingales in survival analysis”, Electronic Journal for History of Probability and Statistics. June, Vol. 5 No.1.

Andersen, P. K., Gill, R. D., (1982), “Cox’s Regression Model for Counting Process: A Large Sample Study”, Annals of Statistics, 10, 1100-1120.

Andersen P. K, Borgan Ø, Gill R, Keiding N., (1993), Statistical Models Based on Counting Processes. Springer-Verlag, NewYork.

Andersen, P. K., Keiding, N., (1998), “Survival Analysis: overview”, In Encyclopedia of Biostatistics (eds P. Armitage and T. Colton), vol.6, 4452-4461. Wiley, Chichester.

Clayton D, Cuzick J., (1985), “Multivariate generalizations of the proportional hazards model”, Journal of the Royal Statistical Society (A), 148: 82-117.

Cox, D. R., (1972), “Regression models and life tables”, Journal of the Royal Statistical Society, B 34, 187-220.

Cox D. R., (1975), “Partial Likelihood”, Biometrika, 62: 269-276.

Fleming, T. R., Harrington, D. P., (2005), Counting Processes and Survival Analysis. Wiley, New York.

Gill, R. D., (1980), “Censoring and stochastic integrals”, Mathematical Centre Tracts, Vol. 124, Mathematisch Centrum, Amsterdam.

Gill, R. D., (2005), “Product-integration. In Encyclopedia of Biostatistics (eds P. Armitageand, T. Colton), vol. 6, 4246-4250.

Gill, R. D., Johansen, S., (1990), “A survey of product-integration with a view towards application in survival analysis”, Annals of Statistics, 18, 1501-1555.

Gustafson, P., (1995), “Hierarchical Bayesian analysis of clustered survival data", Technical Report #144, Department of Statistics, University of British Columbia, Canada.

Hanagal D. D., (2011), Modelling Survival Data Using Frailty Models. Chapman and Hall.

Hougard P., (1984), “Life table methods for heterogeneous populations”, Biometrika, 71: 75-83.

Hougard P., (1986), “Survival models for heterogeneous populations derived from stable distributions”, Biometrika, 73: 387-396.

Huber-Carol C, Vonta F., (2004), “Frailty models for arbitrarily censored and truncated data”, Lifetime Data Analysis, 10: 369-388.

Kaplan, E. L., Meier, P., (1958), “Non-parametric Estimation From Incomplete Observations”, Journal of the American Statistical Association, 53, 457–481.

Lin, D. Y., Ying, Z., (1994), “Semiparametric analysis of the additive risk model”, Biometrika, 81, 61-71.

Mantel N., Haenszel W., (1959), “Statistical aspects of the analysis of data from retrospective studies of disease”, Journal of the National Cancer Inst., 22: 719-748.

McKeague, I. W., Sasieni, P. D., (1994), “A partly parametric additive risk model”, Biometrika, 81, 501-514.

Nelson, W., (1972), “Theory and applications of hazard plotting for censored failure data”, Technometrics, 14, 945–965.

Sinha, D., (1993 a), “Semiparametric Bayesian analysis of multiple time data". J. Amer. Statist. Assoc., 88, 979-983.

Tarone E. R, Ware J., (1977), “On distribution-free tests for equality of survival distributions”, Biometrika, 64: 156-160.

Vaupel, J. W., Manton, K. G., Stallard, E., (1979), “The impact of heterogeneity in inddividual frailty on the Dynamics of mortality”, Demography, 16, 439-454.

Wienke A., (2011), Frailty Models in Survival Analysis. Chapman and Hall.

Yashin A. I,, Vaupel J. W, Iachine A. I., (1995), “Correlated ındividual frailty: An advantageous approach to survival analysis of bivariate data”, Mathematical Population Studies, Vol. 5(2): 145-159.

Yau K. K. W., (2001), “Multilevel models for survival analysis with random effects”, Biometrics, 57: 96-102.