

SAĞDAN SANSÜRLÜ VERİLER İÇİN PARAMETRİK REGRESYON MODELİ VE KEMİK İLİĞİ NAKLİNDE KULLANIMI

Yüksel TERZİ*, Yüksel BEK**, Mehmet Ali CENGİZ*

ÖZET

Sağkalım analizi çalışmalarında kullanılan veri tipleri genelde sağdan sansürlü verilerdir. Bu tür çalışmalarda amaç, hastalığın sağkalım süresine etki eden prognostik faktörleri (açıklayıcı değişkenler) belirlemektir. Bunun için Cox regresyon modeli (Cox oransal hazard modeli) kullanılmaktadır. Ancak sağkalım süreleri bazen parametrik bir dağılım gösterebilir. Bu durumda parametrik regresyon modelleri kullanılır.

Bu çalışmada kemik iliği naklinde kullanılan interferon- α tedavisinden elde edilen sağdan sansürlü veriler incelendi. Bu veriler kullanılarak Cox regresyon modeli kuruldu ve interferon- α tedavisine etki yapan önemli prognostik faktörler belirlendi. Önemli olduğu düşünülen değişkenler tabakalı alınarak, tabakalı Cox regresyon modeli kuruldu ve tabakalı Cox regresyon modeli ile tabakasız Cox regresyon modeli sonuçları karşılaştırıldı. Ayrıca sağkalım sürelerine uygun parametrik dağılımlar, çeşitli yöntemlerle (AIC, logL, artıklar ve grafik yöntemi) incelendi ve en uygun dağılımın log-normal dağılım olduğu görüldü. Log-normal regresyon modeli ile tabakalı ve tabakasız Cox regresyon modeli sonuçları karşılaştırıldı.

Anahtar kelimeler: Sağdan sansürlü veri, sağkalım analizi, Cox regresyon modeli, interferon- α tedavisi.

1. GİRİŞ

Sağlık Bilimleri alanındaki çalışmaların çoğu yeni bir ilacın, uygulamanın veya yeni bir yöntemin kullanımının araştırılması veya hastalığın sebeplerinin ortaya konması ve risk faktörlerinin paylarının belirlenmesi ile ilgilidir. Bazı durumlarda da uygulamaların etki süreleri, olumsuzluğun tekrar nüksetme süreleri veya deneklerin yaşam sürelerinin bilinmesi önem kazanmaktadır.

Bu gibi durumlarda uygulamanın etki süresi, olumsuzlukların yinelenmesine etki eden faktörlerin belirlenmesi, hastaların sağkalım sürelerine etki eden prognostik faktörlerin belirlenmesi, kontrol edilmesi ve izlenmesi ile hastanın yaşam süresinin uzatılması amaçlanmaktadır.

* Ondokuz Mayıs Üniversitesi, Fen-Edebiyat Fak. İstatistik Bölümü (yukselt@omu.edu.tr)

** Ondokuz Mayıs Üniversitesi, Tıp Fakültesi Biyoistatistik Anabilim Dalı

Sağlık alanıyla ilgili çalışmalarda belirli bir süre içinde elde edilen verilerde, bireylerin bir kısmı ölüren, bir kısmı da yaşamını sürdürebilir veya çalışmadan çıkabilir. Yaşamını sürdüren, çalışmadan çekilen veya kaybolan bireylere ait veriler sansürlü veri olarak adlandırılır. Sansürlü veriler için sağkalım analizi yöntemleri kullanılır. Sağlık alanında bir çok kronik hastalıkta olduğu gibi, bir hastalığın varlığı yada yokluğu tek başına sağkalım yada ölüm açısından belirleyici değildir. Hastalıktan hariç olarak bir çok faktör sağkalım süresini uzatabilir veya ölümü hızlandırabilir. Sağkalım analizi çalışmalarında zaman içindeki değişimi dikkate alan ve sansürlü verilerle analiz yapmayı kolaylaştıran bir yonteme ihtiyaç duyulmuştur. Bu yontem Cox (1972) tarafından önerilen Cox regresyon modelidir.

Son yıllarda değişik hastalık türlerine göre elde edilen verilere parametrik regresyon modelleri uygulanmıştır. Bunlardan Kardaun (1983) ile Klein ve Moeschberger (1997) kemik iliği nakli verilerinin log-normal dağılım, Sickle-Satanello vd.(1988) erkeklerde gırtlak kanseri verilerinin log-lojistik dağılım, Nahman vd. (1992) böbrek diyaliz hastaları verilerinin Weibull dağılım, Kim ve Ibrahim (2000) akciğer kanseri verilerinin Weibull dağılımlı olduklarını göstermişlerdir.

2. COX REGRESYON MODELİ

Sansürlü verilerin yer aldığı sağkalım analizinde, bağımlı değişken (sağkalım süresi) ile açıklayıcı değişkenler (prognostik faktörler) arasındaki neden-sonuç bağıntısını ortaya koymak için yararlanılan regresyon yöntemine Cox regresyon modeli adı verilir. Bu yöntemde modele dahil edilecek açıklayıcı değişkenler sürekli, kesikli yada her ikisinin karışımı olabilmektedir.

Cox regresyon modelinde açıklayıcı değişkenler arasında oransal ilişkiler bulunmaktadır. Bu yüzden Cox regresyon modeline oransal hazard modeli de denilmektedir. Cox regresyon modelinde açıklayıcı değişkenlerin, bağımlı değişkene etkileri çarpımsaldır. Regresyon katsayılarının etkilerinin ölçümünde doğrusal regresyonda β alınırken, Cox regresyon modelinde hazard oranını ifade eden $\exp(\beta)$ alınır (Kleinbaum, 1996).

Cox regresyon modelinde dört varsayım vardır:

- i) Açıklayıcı değişkenlerin hazard fonksiyonu üzerine etkileri log-lineer olmalıdır.
- ii) Açıklayıcı değişkenlerin log-lineer fonksiyonu ile hazard fonksiyonu arasında çarpımsal bir ilişki vardır.
- iii) Zaman eksenini boyunca her hangi iki bireyin hazard oranı (HR) sabittir.
- iv) Her bireye ait açıklayıcı değişken değerleri, çalışmanın başlangıcında elde edilmiştir ve çalışma süresince değişmemektedir (Kleinbaum, 1996).

Açıklayıcı değişkenler vektörü x ve sağkalım süresi t olsun. Böylece bir bireyin açıklayıcı değişkenlere göre hazard fonksiyonu $h(t;x)$ ile gösterilir. Bu durumda Cox oransal hazard modelinin matematiksel ifadesi aşağıdaki gibidir:

$$h(t;x)=h_0(t)\exp(\beta' x) \quad (1)$$

Bu modelde;

$h(t;x)$: hazard fonksiyonu

- x : $p \times 1$ boyutlu açıklayıcı değişkenler vektörü
 β' : $1 \times p$ boyutlu bilinmeyen regresyon katsayıları vektörü
 $h_0(t)$: temel hazard fonksiyonu

göstermektedir. Regresyon modellerin aksine, Cox regresyon modelinde sabit terim yoktur. Sabit terim yerine $h_0(t)$ terimi modelde yer almaktadır.

Cox regresyon modeli yarı-parametrik bir modeldir. Bu modelde $h_0(t)$ ve β katsayıları bilinmeyenler, açıklayıcı değişkenler ise bilinenleri göstermektedir. Ayrıca temel hazard fonksiyonu $h_0(t)$ 'nin biçimine ilişkin parametrik bir varsayım yapılmaması ve açıklayıcı değişkenlerin hazard fonksiyonu üzerindeki etkisinin ise parametrik olarak ifade edilmesi, bu modelin yarı-parametrik olduğu anlamına gelmektedir.

Cox regresyon modelinde parametre tahminleri kısmi olabilirlik fonksiyonu ile bulunabilir (Klein ve Moeschberger, 1997).

$$L(\beta) = \prod_{i=1}^k \frac{\exp(\beta'x_i)}{\sum_{j \in R(t_i)} \exp(\beta'x_j)} \quad (2)$$

Tahmin edilen parametrelerin anlamlılıkları ise Wald, Olabilirlik veya Skor testi ile test edilir.

Kurulacak modelin istatistiksel olarak anlamlı olup olmadığı, olabilirlik oran testi ile test edilebilir. Bunun için her ilave değişken için $H_0 : \beta = 0$ olan yokluk hipotezi test edilir.

$$LR = -2 \ln \left[\frac{L_0}{L_i} \right] \sim \chi_p^2 \quad (3)$$

(3) eşitliğinde L_0 ortak deęişkensiz modelin olabilirlik deęerini, L_i ise ortak deęişkenli modelin olabilirlik deęerini göstermektedir. Bulunan olabilirlik oran testi deęeri (LR), p serbestlik dereceli ki-kare daęılımını gösterir. Eđer hipotez reddedilirse ($\beta_i \neq 0$), en az bir regresyon katsayısı sıfırdan farklıdır yani kurulan model geęerlidir denilir.

Modelin önemlilik testi yapıldıktan sonra, elde edilen olabilirlik fonksiyonuna en yüksek katkılı deęişkenleri modele alarak, yeni bir model kurmak (azaltılmıř model) ve bu modeli tüm deęişkenleri kapsayan tam (full) modelle karřılařtırmak gerekir. Bu işlemler olabilirlik oran testi ile yapılabilir. Yokluk hipotezi ve test istatistięi ařaęıdaki gibi kurulur:

H_0 : Modelden çıkarılan deęişkenlere ait $\beta_i = 0$, $i=1,2,\dots,p$

$$LR = -2 \ln \left[\frac{L_1}{L_2} \right] \sim \chi_{(p_1-p_2)}^2 \quad (4)$$

(4) eşitliğinde L_1 tam modelin olabilirlik deęerini, L_2 ise indirgenmiř modelin olabilirlik deęerini göstermektedir. Burada p_1 tüm deęişkenleri içeren modeldeki deęişken sayısı, p_2 ise indirgenmiř modeldeki deęişken sayısıdır. Eđer hipotez reddedilirse, deęişken

sayısı azaltılmış modelin, tam model kadar iyi olduğu söylenebilir. Bir başka deyişle indirgenmiş model dışında kalan değişkenlerin olabilirlik fonksiyonuna katkısı önemsizdir. Hipotez kabul edilirse, modelden çıkan değişkenin modele katkısı olmadığı anlaşılır.

Oransal hazard modeli için, hazard oranının tüm zaman boyunca sabit olması gerekir. Bir bireyin hazardı diğer bireyin hazardı ile orantılıdır. Bu orantı ise zamandan bağımsızdır.

$$\hat{HR} = \frac{\hat{h}(t, x^*)}{\hat{h}(t, x)} = \exp \left[\sum_{i=1}^p \hat{\beta}_i (x_i^* - x_i) \right] = \hat{\theta} \quad (5)$$

Bir açıklayıcı değişkenin iki veya daha çok kategorisi için hazard oranları keşiyorsa oransal hazard varsayımı gerçekleşmez (Kleinbaum, 1996).

Oransal hazard varsayımını sağlayan değişkenler Cox regresyon modeline dahil edilirken, bu varsayımı sağlamayan değişkenler ise modele katılmazlar. Oransal hazard varsayımını sağlamayan bir açıklayıcı değişken tabakalar haline getirilerek, Cox regresyon modeli yeniden düzenlenir.

z_1, \dots, z_k oransal hazard varsayımını sağlamayan değişkenler ve x_1, \dots, x_p ise oransal hazard varsayımını sağlayan değişkenler olsun. Tabakalı Cox regresyon modelini uygulamak için, önce oransal hazard varsayımını sağlamayan ve etkisi önemli olduğu düşünülen bir z^* değişkeni tabakalara ayrılır ve her bir kategori bir tabaka olarak alınır. Bu durumda tabakalı Cox regresyon modeli aşağıdaki gibi yazılabilir:

$$h_g(t, x) = h_{0g}(t) \exp [\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p] \quad , \quad g=1, 2, \dots, k^* \quad (6)$$

Bu modelde g, z^* ile tanımlanan tabakalı değişkenlerin indisini göstermektedir. Oransal hazard varsayımını sağlayan değişkenler (x_1, \dots, x_p) modelde yer alırken, oransal hazard varsayımını sağlamayan değişkenler (z_1, \dots, z_k) ise modelde yer almamaktadır. Ayrıca her bir tabaka için temel hazard fonksiyonları [$h_{0g}(t)$] farklı iken, $\beta_1, \beta_2, \dots, \beta_p$ regresyon katsayıları aynıdır (Kleinbaum, 1996).

2.1. Cox Regresyon Modelinde Artıklar

Literatürde artıklarla ilgili olarak sunulan ilk artık yöntemi Cox-Snell (1968) artıklarıdır. Cox-Snell artıkları Cox regresyon modelinin uyum iyiliğini ölçmek için kullanılır. Bu artıklar aşağıdaki gibi tanımlanır:

$$r_j = \hat{H}_0(t_j) \theta_j \quad , \quad j = 1, \dots, n \quad (7)$$

Burada θ_j

$$\theta_j = \exp \left(\sum_{i=1}^p \beta_i x_{ji} \right) \quad , \quad j = 1, 2, \dots, n$$

olarak tanımlanır. Bu ifadede β 'lar tahmin edilen regresyon katsayılarını göstermektedir. (7) eşitliğinde $\hat{H}_0(t_j)$ birikimli temel hazard fonksiyonunun tahmin edicisidir (Klein ve Moeschberger, 1997).

$$\hat{H}_0(t_j) = \sum_{t_i \leq t_j} \frac{d_i}{\sum_{r \in R(t_i)} \theta_r}$$

Burada d_i , t_i zamanındaki ölen bireylerin sayısını, $R(t_i)$ ise risk altındaki birey sayısını göstermektedir.

3. PARAMETRİK REGRESYON MODELİ

Sağkalım zamanı verilerininin logaritması ile ortak değişkenlerin değerleri arasındaki ilişki doğrusaldır. Sağkalım zamanının logaritması alınarak ortak değişkenlerin yer aldığı doğrusal model aşağıdaki gibi kurulur.

$$y = \ln(t) = \mu + \beta'x + \sigma W \quad (8)$$

Bu modelde W hata dağılımı değişkenidir. Sansürlü verilerde W rasgele değişkeni için çeşitli modeller kurulabilir. Mesela W 'nin dağılımı Weibull modelinde standart uç değer dağılımı, log-lojistik dağılımda standart lojistik dağılım, log-normal dağılımda ise standart normal dağılım gösterir.

Seçilen parametrik modelin uygun olup olmadığı grafik yöntemi, artıklar, logL değerleri ve Akaike bilgi kriteri (AIC) ile kontrol edilebilir. Parametrik regresyon modellerinden Weibull regresyon modeli için hazard fonksiyonu ve sağkalım fonksiyonu aşağıdaki gibidir (Klein ve Moeschberger, 1997):

$$h(t; x) = h_0(t) \exp(\beta'x) = \alpha \lambda (\lambda t)^{\alpha-1} \exp(\beta'x) \quad (9)$$

$$S(t; x) = \{ \exp[-(\lambda t)^\alpha] \}^{\exp(\beta'x)} \quad (10)$$

4. UYGULAMA

Kemik iliği nakli akut lösemili hastalar için standart bir tedavidir. Lösemiler için kemik iliği naklinde interferon- α tedavisi kemoterapiden daha iyi sonuç verir. Bunun için kemik iliği naklinde interferon- α tedavisi önerilmiştir. İnterferon iğne ile verilen bir ilaç tedavisidir. Konu ile ilgili olarak, Temmuz 1991 ile Temmuz 1999 yılları arasında Ankara Ün. İbni Sina Has., İstanbul Ün. Tıp. Fak. ve GATA Kemik İliği Nakli Bölümlerinden alınan 106 hastaya ait veriler aylık incelenmiştir. Hastalardan 48 tanesi ölü, 58 tanesi ise sansürlü gözlem olarak incelenmiştir.

Önce kesikli ve sürekli açıklayıcı değişkenler ayrı ayrı test edilip, oransal hazard varsayımını sağlayıp sağlamadığı incelendi. Sonuçta interferon- α (x_1), genetik bozukluk (x_2), kronik gvhd (x_3), nakil sonrası durum (x_4), nakil esnasında ölüm (x_5), risk (x_6), ölüm nedenleri (x_7), yaşkod* x_7 , x_3 * x_7 ve x_5 * x_7 değişkenleri önemli bulunup, modele alınmışlardır. Burada sürekli bir açıklayıcı değişken olan yaş önemsiz bulunmuştur. Ancak önemli olduğu düşünülen bu değişken uygun biçimde kodlanarak yeniden incelenmiş ve modele katılımı sağlanmıştır. Tablo 1'de Cox regresyon modeli sonuçları verilmiştir.

Tablo 1. Cox regresyon modeli sonucu

Değişkenler	B	SE	Wald	sd	P	Exp(B)	Exp(B) için güven aralığı	
							Alt sınır	Üst sınır
x ₁	-.724	.404	3.203	1	.073	.485	.220	1.071
x ₃	-2.828	.944	8.971	1	.003	.059	.009	.376
x ₄	.940	.346	7.386	1	.007	2.559	1.300	5.041
x ₅	6.411	1.666	14.804	1	.000	608.370	23.223	15937.471
yaşkod*x ₇	.879	.143	37.729	1	.000	2.407	1.819	3.186
x ₅ *x ₇	-1.662	.517	10.325	1	.001	.190	.069	.523

Kemik iliği naklinde cinsiyet uygunluğu ve transplant tipi açıklayıcı değişkenleri tabaka olarak alınarak, tabakalı Cox regresyon modeli yapılmış ve Tablo 1'deki aynı sonuçlar elde edilmiştir. Transplant yılı (<1995, ≥1995) tabaka olarak alındığında ise Tablo 1'den farklı sonuç bulunmuştur.

Tablo 2. Transplant yılı değişkeninin tabakalı alındığı Cox regresyon modeli sonucu

Değişkenler	B	SE	Wald	sd	P	Exp(B)	Exp(B) için güven aralığı	
							Alt sınır	Üst sınır
x ₃	-2,840	,980	8,397	1	,004	,058	,009	,399
x ₄	,978	,374	6,850	1	,009	2,659	1,278	5,532
x ₅	5,895	1,542	14,605	1	,000	363,157	17,666	7465,337
yaşkod*x ₇	,882	,150	34,605	1	,000	2,416	1,801	3,242
x ₅ *x ₇	-1,642	,493	11,067	1	,001	,194	,074	,509

Verilerin parametrik bir dağılım gösterip göstermediği araştırılmıştır. Sağkalım analizi çalışmalarında önemli bir yer tutan bazı dağılımlara göre verilerin uygunluğuna bakılmıştır. Elde edilen sonuçlar Tablo 2'de verilmiştir.

Tablo 3. Bazı dağılımların logL ve AIC değerleri

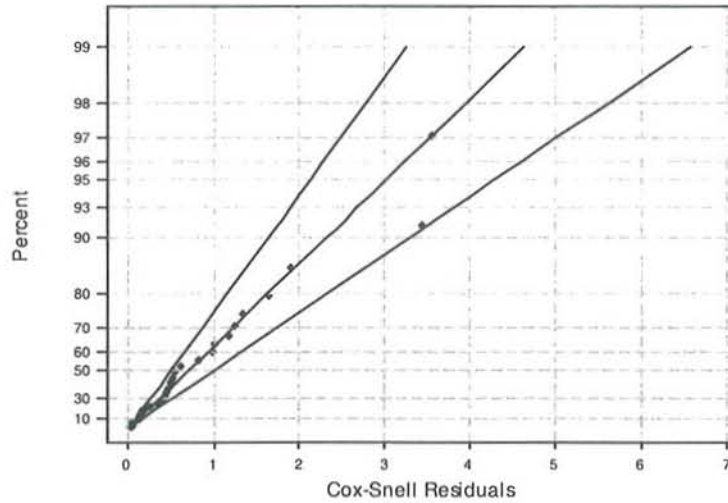
Dağılımlar	logL	AIC
Üstel	-60.430067	138,860
Weibull	-58.392668	136,785
Log-lojistik	-56.767342	133,534
Log-normal	-56.728226	133,456

Tablo 3'den interferon- α verisi için en uygun dağılımın log-normal dağılım olduğu tespit edilmiştir. Böylece log-normal dağılıma göre regresyon modeli yeniden kurulmuş ve elde edilen sonuçlar Tablo 4'de gösterilmiştir.

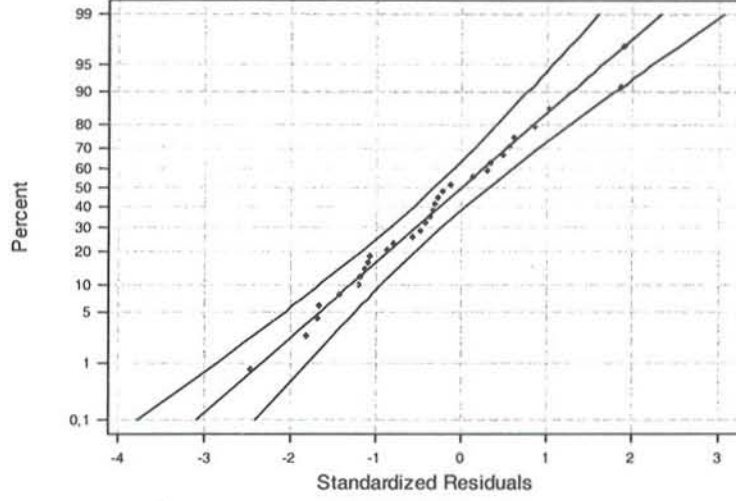
Tablo 4. Log-normal dağılıma göre regresyon modeli sonucu

Değişkenler	B	S.E.	Z	P	Güven Aralığı	
Kesen(μ)	1.4811	0.2995	4.94	0.000	0.8940	2.0682
x_1	0.1723	0.1064	1.62	0.105	-0.0363	0.3808
yaşkod	0.0652	0.1049	0.62	0.534	-0.1404	0.2709
x_2	0.1039	0.0841	1.24	0.217	-0.0609	0.2688
x_3	0.8101	0.2552	3.17	0.001	0.3100	1.3103
x_4	-0.2722	0.0919	-2.96	0.003	-0.4523	-0.0921
x_5	-1.3312	0.2486	-5.35	0.000	-1.8184	-0.8440
x_6	0.0110	0.0599	0.18	0.853	-0.1063	0.1284
x_7	-0.1617	0.1675	-0.97	0.335	-0.4901	0.1667
yaşkod* x_7	-0.1923	0.0754	-2.55	0.011	-0.3403	-0.0444
x_3 * x_7	-0.1449	0.1843	-0.79	0.432	-0.5061	0.2163
x_5 * x_7	0.5266	0.1410	3.73	0.000	0.2502	0.8030
Ölçek(σ)	0.3164	0.0419			0.2439	0.4102

Log-normal regresyon modelinin uygunluğu için Cox-Snell ve standardize edilmiş artıklar Şekil 1 ve Şekil 2'de verilmiştir. Bu iki artık eğrilerinde bir sapma olmadığından, bulunan log-normal modelin uygun olduğu söylenebilir.

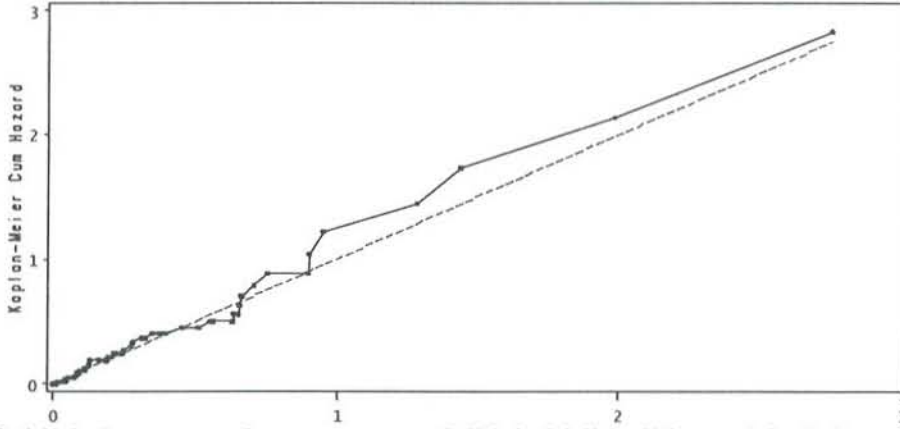


Şekil 1. Log-normal regresyon modeli için Cox-Snell artıkları



Şekil 2. Log-normal regresyon modeli için standardize edilmiş artıklar

Parametrik regresyon modellerinin uygunluğu grafik yöntemiyle de görülebilir. Log-normal dağılımlı regresyon modeli için grafik sonucu Şekil 3'de verilmiştir.



Şekil 3. Log-normal regresyon modeli için birikimli hazard fonksiyonu eğrisi

5. SONUÇ

Kemik iliği naklinde kullanılan interferon- α tedavisi ile ilgili çalışmada elde edilen veriler sağdan sansürlü verilerdir. Sağdan sansürlü bu veriler kullanılarak kurulan Cox regresyon modelinde, Tablo 1'de görüldüğü gibi x_1 (interferon- α), x_3 (kronik gvhd), x_4 (nakil sonrası durum), x_5 (nakil esnasında ölüm) değişkenleri ve yaşkod* x_7 ile x_5 * x_7 etkileşimleri önemli bulunmuştur. Tablo 1'de önemli bulunan x_1 değişkeninin hazard oranı [$\exp(B)=0,485$] birden küçük çıkmış olup, bu ifade bize interferon alan hastaların ölüm riski almayanlara göre %51,5 daha az demektir. Diğer değişkenlerde benzer biçimde yorumlanabilir.

“Cinsiyet uygunluğu” ve “transplant tipi” açıklayıcı değişkenleri tabakalı olarak alınarak, tabakalı Cox regresyon modeli yapıldı ve Tablo 1’deki sonuçlar elde edildi. Bu iki açıklayıcı değişkenin oransal hazard varsayımını sağlamaması bu sonucun elde edilmesine neden olmuştur.

“Transplant yılı” tabaka alındığında ise Tablo 1’de önemli bulunana açıklayıcı değişkenlerden x_1 değişkeni önemsiz bulunmuştur. Bunun nedeni incelendiğinde ise, transplant yılının oransal hazard varsayımını sağladığı ve 1995 yılından küçük olan yıllarda interferon- α tedavisinin farklı sonuç gösterdiği görüldü.

Log-normal regresyon modelinde ise, Cox regresyon modelinde önemli bulunan x_1 (interferon- α) değişkeni önemsiz bulunmuştur. Yani log-normal regresyon modeli ile transplant yılının tabakalı alındığı tabakalı Cox regresyon modelinde aynı değişkenler önemli bulunmuştur. Tablo 3’de önemli bulunan x_4 (nakil sonrası durum) değişkeninin regresyon katsayısı -0,2722 bulunmuştur. Bu ifade bize Akut, Kronik ve Akut+Kronik aşamalarındaki hastalar, nakil sonrası bu belirtilerin olmadığı hastalara göre daha az yaşadığını göstermektedir. Diğer önemli bulunan değişkenler benzer biçimde yorumlanabilir.

Bu çalışmaya öneri olarak parametre tahminlerinin Bayesci yaklaşımla bulunması ve yine Bayesci yaklaşım yeni model kurulumu yapılarak, sonuçlar karşılaştırılabilir. Ayrıca değişik hastalık türleriyle yeni parametrik modeller kurularak, tabakalı ve tabakasız Cox regresyon modelleriyle sonuçlar tartışılabilir.

KAYNAKLAR

- COX, D.R. (1972), Regression models and life tables, Journal of the Royal Statistical Society, 34, 187-220.
- COX, D.R. and Snell, E.J. (1968), A general definition of residuals (with discussion), Journal of the Royal Statistical Society, B, 30, 248-275.
- HOSMER, D.W. and Lemeshow, S. (1999), Applied Survival Analysis: Regression Modeling of Time to Event Data, Wiley, John and Sons, Erişim: [http://www.ats.ucla.edu/stat/sas/examples/asa/default.htm]. Erişim Tarihi: 25.09.2003
- KARDAUN, O. (1983), Statistical Analysis of Male Larynx-Cancer Patients-A case Study, Statistical Nederlandica, 37, 103-126.
- KIM, S.W. and IBRAHİM, J.G. (2000), On Bayesian Inference for Proportional Hazards Models Using Noninformative Priors, Lifetime Data Analysis, 6, 331-341.
- KLEIN, J.P. and MOESCHBERGER, M. L. (1997), Survival Analysis-Techniques for Censored and Truncated Data, Springer.
- KLEINBAUM, D.G. (1996), Survival Analysis, A Self Learning Text, Springer, New York.
- NAHMAN, N.S., MIDDENDORF, D.F., BAY, W.H., MCELLIGOTT, R., POWELL, S. and ANDERSON, J. (1992), Modification of the Percutaneous Approach to Peritoneal Dialysis Catheter Placement Under Peritoneoscopic Visualization: Clinical Results in 78 Patients, Journal of the American Society of Nephrology 3, 103-107.

Sickle-Santanello, B.J., Farrar, W.B., Keyhani-Rofagha, S., Klein, J.P., Pearl, D., Laufman, H., Dobson, J. and O'Toole, R.V. (1988), A Reproducible System of Flow Cytometric DNA Analysis of Paraffin Embedded Solid Tumors: Technical Improvements and Statistical Analysis, *Cytometry* 9, 594-599.

EK

Uygulamada elde edilen sonuçların kullanıldığı SAS macro programı (Hosmer ve Lemeshow, 1999):

title 'model building';

```
proc phreg data=ifn;
  model dfs*event(0)=yaskod x1 x2 x3 x4 x5 x6 x7 int1 int2 int3;
  int1=yaskod*x7; int2=x3*x7; int3=x5*x7; run;
```

title 'proportionality assumption';

```
proc phreg data=ifn;
  model dfs*event(0)=x1 x3 x4 x5 int1 int3 t1 t3 t4 t5 gint1 gint3;
  t1=x1*log(dfs); t3=x3*log(dfs); t4=x4*log(dfs); t5=x5*log(dfs);
  int1=yaskod*x7; int3=x5*x7;
  gint1=int1*log(dfs); gint3=int3*log(dfs);
  proportionality_test: test t1, t3, t4, t5, gint1, gint3; run;
```

title 'Cox-Snell residuals';

```
proc phreg data = ifn;
  model dfs*event(0) =yaskod x1 x2 x3 x4 x5 x6 x7 int1 int2 int3;
  int1=yaskod*x7; int2=x3*x7; int3=x5*x7;
  output out = figure1_1 LOGSURV = h; /*-logsurv is the cox-snell residual*/ run;
data figure1_1a;
  set figure1_1; h = -h; cons = 1; run;
proc phreg data = figure1_1a ;
  model h*event(0) = cons; output out = figure1_1b logsurv = ls /method = ch; run;
  data figure1_1c; set figure1_1b; haz = - ls; run;
proc sort data = figure1_1c; by h; run;
title "Cox-Snell Residuals";
  axis1 order = (0 to 3 by .5) minor = none;
  axis2 order = (0 to 3 by .5) minor = none label = ( a=90);
  symbol1 i = stepjl c= blue; symbol2 i = join c = red l = 3;
proc gplot data = figure1_1c;
plot haz*h =1 h*h =2 /overlay haxis=axis1 vaxis= axis2;
label haz = "Estimated Cumulative Hazard Rates"; label h = "Residual"; run; quit;
```

title 'Parametric regression models';

```
ods output ParameterEstimates (match_all=bydist persist=proc) = ifn1;
proc lifereg data = ifn;
  model dfs*event(0)=yaskod x1 x2 x3 x4 x5 x6 x7/distribution=exponential; run;
proc lifereg data = ifn;
  model dfs*event(0)= yaskod x1 x2 x3 x4 x5 x6 x7 /distribution=weibull;run;
```

```

proc lifereg data = ifn;
model dfs*event(0)= yaskod x1 x2 x3 x4 x5 x6 x7 /distribution=llogistic; run;
proc lifereg data = ifn;
    model dfs*event(0)= yaskod x1 x2 x3 x4 x5 x6 x7 / distribution=lognormal;run;
ods output close;
    data ifn1; set &bydist; retain group 0; keep parameter estimate stderr group;
    if parameter = "Intercept" then group= group+1; run;
proc format;
    value group    1= "Exponential"  2 = "Weibull"
                  3 = "Log Logistic" 4 = "Log Normal"
run; options nocenter formchar='|----|+|----+=|~^<>*';
proc tabulate data=ifn1;
    format group group.; class group parameter; var estimate stderr;
    table parameter=, group=' *(estimate='Est' stderr='SE')*sum=' *f=5.3/RTS=13; run;
ods output close;
    data ifnbtm; set &bydist; lagv = lag(cvalue1);
    if label1 = "log likelihood" or label1="name of distribution";
    keep nvalue1 aic; if lagv = "Exponential" then aic = -2*nvalue1 + 2*(8 +1);
    else aic = -2*nvalue1 + 2*(8+1+1); if nvalue1 ~.=.;
run; data ifnbtm1; set ifnbtm; id = _n_; run;
proc format;
    value group    1 = "Exponential"  2 = "Weibull"
                  3 = "Log Logistic" 4 = "Log Normal" run;
proc print data=ifnbtm1(rename=(nvalue1=LogL)) noobs;
    format id group.; var id LogL AIC; run;

title 'Cox-Snell residual plot for log normal regression model';
proc reliability data =ifn1;
    distribution lognormal; model dfs*event(0) = yaskod x1 x2 x3 x4 x5 x6 x7
        /obstats (survival);
ods output ModObstats = figure1_1; run;
    data figure1_1a; set figure1_1; h = -log(surv); run;
proc phreg data = figure1_1a ;
    model h*event(0) = ; output out = figure1_1b logsurv = ls /method = ch; run;
    data figure1_1c; set figure1_1b; haz = - ls; run;
proc sort data = figure1_1c; by h; run;
    axis1 order = (0 to 2 by .5) minor = none;
    axis2 order = (0 to 2 by .5) minor = none label = ( a=90);
    symbol1 i = i=1|p c= blue; symbol2 i = join c = red l = 2;
proc gplot data = figure1_1c;
    plot haz*h =1 h*h =1 /overlay haxis=axis1 vaxis= axis2;
    label haz = "Estimated Cumulative Hazard Rates"; label h = "Residual";
run; quit;

```

PARAMETRIC REGRESSION MODEL FOR RIGHT CENSORED DATA WITH APPLICATION TO BONE MARROW TRANSPLANT

ABSTRACT

The data type used in survival analysis is generally right censored data. The aim in this kind of studies is to determine prognostic factors which have effects on survival time. For the above reason Cox regression model is used. Nevertheless survival times sometimes may have parametric distribution. In this case parametric regression models are used.

In this study, the right censored data obtained from interferon- α treatment used in bone marrow transplant were investigated. Using this data, Cox regression model was constructed and prognostic factors, which have effects on interferon- α treatment, were determined. Taking variables which are thought important, as stratified, stratified Cox regression model was constructed. Stratified Cox regression model and Cox regression model was compared. Also parametric distributions, which a suitable for survival times, were examined using various methods such as AIC, LogL, residuals and graphic method. Log-normal regression model was selected as the most appropriate model. Latter model and the other models examined in this study are compared.

Key Words: *Right censored data, survival analysis, Cox regression model, interferon- α treatment*