



Sistolik Kan Basıncının Tahmininde Yanlı Regresyon Yöntemlerinin Kullanılması⁺

Esra Pamukçu*, Cemil Çolak**, Sinan Çalık*, Zülfiye Kuzu***

* Fırat Üniversitesi Fen Fakültesi, İstatistik Bölümü, Elazığ

** İnönü Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Malatya

*** Fırat Üniversitesi Tıp Fakültesi, Kardiyoloji Anabilim Dalı, Elazığ

Amaç: Bu çalışmada, sistolik kan basıncının (SKB) tahmin edilmesinde En Küçük Kareler Regresyonu (LS), Ridge Regresyonu (RR) ve Temel Bileşenler Regresyonu (PCR) yöntemlerine ilişkin sonuçların incelenmesi amaçlandı.

Gereç ve Yöntemler: Bu çalışma kesitsel ve analitik bir çalışma olup, karşılaştırma amacıyla herhangi bir kontrol grubu oluşturulmamıştır. Yaş, boy, karın çevresi, glikoz, HDL, LDL, üre, kreatinin, potasyum, HGB, HCT ve diastolik kan basıncı açıklayıcı değişkenler olarak ele alınmıştır. Veriler 1 Ekim 2009 ile 22 Aralık 2009 tarihleri arasında Fırat Üniversitesi Tıp Merkezi Kardiyoloji Polikliniğine gelen 128 hastadan elde edilmiştir.

Bulgular: LS, RR ve PCR için belirlilik katsayıları sırasıyla 0.6003, 0.5876 ve 0.5994; hata kareler ortalamaları sırasıyla 13.305, 13.515 ve 13.319 olarak elde edilmiştir. RR yönteminde k Ridge parametresi; çoklu bağlantının giderildiği ve kuramsal beklentilere uygun sonuçlarının elde edilebildiği ilk değer olan 0.02 seçilmiştir.

Sonuç: Diastolik kan basıncı, karın çevresi ve kreatinin değişkenlerinin, sistolik kan basıncının tahmininde kullanılabileceğine, PCR ve RR yöntemleri ile daha tutarlı ve durağan kuramsal beklentilere uygun tahminler elde edildiği sonucuna varıldı. PCR ve RR sonuçları incelendiğinde, PCR'nin RR'ye göre biraz daha iyi tahminler sağladığı belirlendi.

Anahtar Kelimeler: Çoklu Bağlantı; Sistolik Kan Basıncı; En Küçük Kareler; Ridge Regresyon; Temel Bileşenler Regresyonu.

The Use of Biased Regression Methods in Predicting Systolic Blood Pressure

Objective: In this study, it is aimed to investigate the results of Least Square Regression (LS), Ridge Regression (RR) and Principal Components Regression (PCR) methods in predicting systolic blood pressure.

Material and Methods: This study is cross-sectional and analytical. It had not been occurred a control group for the purpose of comparison. Age, height, abdominal circumference, glucose, HDL, LDL, urea, creatinine, potassium, HGB, HCT and diastolic blood pressure were discussed as explanatory variables. Data were obtained from 128 patients who came to Fırat University Medical Center's Cardiology Clinic between 1 October 2009 and 28 December 2009.

Results: R-square values for LS, RR and PCR were obtained as 0.6003, 0.5876 and 0.5994, respectively. Mean Square Error (MSE) values were obtained as 13.305, 13.515 and 13.319, respectively. Because multicollinearity was resolved and the results accordance with theoretical expectations was obtained, Ridge parameter $k = 0.02$ was chosen in the RR method.

Conclusion: It was decided that diastolic blood pressure, abdominal circumference and creatinine variables can be used to predict systolic blood pressure and a more consistent, more stable and accordance with theoretical expectations estimates had been obtained with RR and PCR methods. According to the results of PCR and RR, it was determined that PCR provides better estimates than RR.

Key Words: Multicollinearity; Systolic Blood Pressure; Least Square; Ridge Regression; Principal Components Regression.

⁺Bu çalışmanın sözlü bir sunumu; Van Yüzüncü Yıl Üniversitesi XII. Biyoistatistik kongresine gönderilmiş ve özeti bildiri kitapçığında basılmıştır.

Giriş

Hipertansiyon, normal düzeyin üzerine çıkan kan

Başvuru Tarihi: 05.08.2010, Kabul Tarihi: 03.11.2010

basıncı seviyesi olarak tanımlanır. Hipertansiyon erken tanı ile kontrol altına alınabilen; geç kalındığında kalp, beyin, böbrekler ve gözlerde hasara neden olabilen ciddi bir hastalıktır.¹ Dünya Sağlık Örgütü (WHO) tahminlerine göre kan basıncı kontrolünün yeterli

düzeyde sağlanamaması, inme vakalarının üçte ikisinden ve iskemik kalp hastalıklarının da yarısından sorumludur.² Çok sayıda epidemiyolojik çalışmada, yüksek kan basıncının kadın ve erkeklerde koroner kalp hastalığı, kalp yetersizliği açısından en önemli risk faktörlerinden biri olduğu saptanmıştır.³ Özellikle 55 yaş üstündeki insanlarda sistolik kan basıncı diğer risk faktörlerine göre daha fazla hipertansiyon ve kardiyovasküler risk belirleyicisidir. Bu yüzden tedavilerde ilk amaç; sistolik kan basıncını 140 mmHg veya daha altına çekmek olmalıdır.⁴ Bu nedenle söz konusu çalışmada, sistolik kan basıncını etkileyebilecek faktörlerin tespiti ve tahmini için yanıt değişkeni olarak sistolik kan basıncı seçilmiştir. Bu çalışmada, sistolik kan basıncının (SKB) tahminlenmesinde En Küçük Kareler Regresyonu (LS, Least Square Regression), Ridge Regresyonu (RR, Ridge Regression) ve Temel Bileşenler Regresyonu (PCR, Principal Components Regression) yöntemlerinin sonuçlarının incelenmesi amaçlanmıştır.

Gereç ve Yöntemler

Çoklu Bağlantı Problemi: Çoklu doğrusal regresyon modelinde açıklayıcı değişkenlerin birbirleriyle ilişkili olmaması varsayımının bozulması çoklu doğrusal bağlantı problemini ortaya çıkarmaktadır. Çoklu bağlantı $X; n \times (k+1)$ boyutlu veri matrisini göstermek üzere, X_1, X_2, \dots, X_k kolonlarının doğrusal bağımsızlığı olarak da tanımlanabilir. O halde lineer (doğrusal) bağımlılık tanımı gereğince t_1, t_2, \dots, t_k en az biri sıfırdan farklı olan skalerler olmak üzere; $\sum_{j=1}^k X_j t_j = 0$ ise tam çoklu bağlantıdan, $\sum_{j=1}^k X_j t_j \cong 0$ ise güçlü çoklu bağlantıdan söz edilir. Bu durumda regresyon çözümlemesi için $(X'X)^{-1}$ hesaplanabilecek; ancak, bu durumun regresyon sonuçları üzerinde olumsuz etkileri olacaktır.⁵

En Küçük Kareler Yöntemi: LS, sürekli ölçümle belirtilen bir yanıt değişkeni ile açıklayıcı değişken veya değişkenler arasında doğrusal bir ilişki mevcut olduğunda, gözlem noktaları arasından geçen en uygun regresyon doğrusunu yani Y gözlem değerine en yakın tahmini \hat{Y} değerini minimum hata ile veren doğrusal fonksiyonu bulmak için kullanılan bir yöntemdir. Çoklu doğrusal regresyonda $Y; n \times 1$ boyutlu bağımlı değişken vektörü, $X; n \times (k+1)$ boyutlu veri matrisi, $\beta; (k+1) \times 1$ boyutlu katsayılar vektörü olmak üzere

$$\beta = (X'X)^{-1}X'Y$$

LS tahmin edicisidir. LS'nin en iyi tahmin edici olabilmesi için bazı varsayımların sağlanması gerektiği unutulmamalıdır.⁶⁻⁸

Ridge Regresyon: RR, çoklu doğrusal regresyon analizinde çoklu bağlantı probleminin ortaya çıktığı durumlarda, veride hiçbir değişiklik yapmadan, veri ekleyip çıkarmadan kullanılan yanlı tahmin tekniklerinden biridir. Hoerl ve Kennard (1970 a,b) tarafından önerilen bu yöntemin, genel olarak varyans ve kovaryans matrisinin köşegen değerlerine küçük bir yanlılık sabiti k ilave etmenin dışında LS ile işleyişi aynıdır. Aşağıda Ridge tahmin edicisi verilmektedir.⁹⁻¹¹

$$\hat{\beta}^* = (X'X + kI)^{-1}X'Y$$

Temel Bileşenler Regresyonu: PCR, her doğrusal regresyon modelinin bir dik açıklayıcı değişkenler kümesine dayanarak yeniden açıklanacağı gerçeği üzerine oluşturulmuş olup, açıklayıcı değişkenler arasında çoklu bağlantı olduğu durumlarda uygulanmaktadır. PCR analizinde açıklayıcı değişkenlere ait veriler standartize edildikten sonra korelasyon matrisinden bulunan standartlaştırılmış asıl özdeğerler ve özdeğerlerden standartlaştırılmış özvektörler belirlenir. Çoklu bağlantının varlığı sebebiyle standartlaştırılmış asıl özvektörlerden sıfıra yakın olanlar anlamlı sonuçlar vermez. Bu nedenle bu özdeğerlere ilişkin özvektörlerin devre dışı bırakılması gerekir. Dolayısıyla ilk olarak önemli özdeğer sayısına karar vermek gerekmektedir. Bu amaçla geliştirilen yöntemlerden en çok bilinenleri; standartlaştırılmış veri matrisinin kullanıldığı durumlarda birden büyük özdeğerlerin sayısını almaktır. Yaklaşık aynı mantığa dayanan

$$\sum_{j=1}^m \frac{\lambda_j}{p} \geq \frac{2}{3}$$

koşulunun sağlandığı en küçük m değeri de önemli temel bileşen sayısını belirlemektedir.¹²⁻¹³

Veriler

Bu çalışma hipertansif hastalık hikâyesine sahip olan hastaların sistolik kan basıncı üzerinde etkili olabileceği düşünülen değişkenlerin etkilerinin değerlendirilmesine yönelik kesitsel ve analitik bir araştırmadır. Veriler İleriye yönelik olarak 01-Ekim-2009/28-Aralık-2009 tarihleri arasında Fırat Üniversitesi Tıp Fakültesi Kardiyoloji Polikliniğine gelen hastalardan toplanan bilgiler ile oluşturulmuştur. Karşılaştırma amacıyla herhangi bir kontrol grubu oluşturulmamış, bağımlı ilişkili olabilecek değişkenler arasında çoklu doğrusal bağlantılar aranmıştır. Hair ve arkadaşlarının (1998) ve Green'in (1991)'de çoklu regresyon analizi için örneklem büyüklüğü ile ilgili görüşleri ve imkânlar göz önünde bulundurulmuş ve $n=128$ olarak belirlenmiştir.¹⁴

Sistolik Kan Basıncının Tahmininde Yanlı Regresyon Yöntemlerinin Kullanılması

Bulgular

Hastaların %50'si kadın, %50'si erkektir. Tüm değişkenlerin normal dağılıma uygunluğu Kolmogorov-Smirnov testi ile incelenmiştir. Normal dağılıma uygun olduğu belirlenen tüm değişkenlere, bağımsız

örneklerde t-testi uygulanması sonucunda cinsiyete göre ortalamalar arasındaki farkların sonuçları ve değişkenlere ilişkin tanımlayıcı istatistikler Tablo 1'de verilmiştir.

Tablo 1. Değişkenlere ait tanımlayıcı istatistikler

Değişkenler	n	Minimum	Maksimum	Ortalama±SD		P
				Kadın	Erkek	
SKB	128	100.00	190.00	141.41±20.92	139.38±19.20	0.568
Yaş	128	20	80	56.34±11.37	59.63±10.31	0.090
Boy	128	1.43	1.90	1.56±0.061	1.71±0.078	<0.001
Karın çevresi	128	0.75	1.30	1.03±0.13	1.01±0.07	0.155
Glikoz	128	15.00	311.00	113.92±43.28	108.81±46.69	0.522
HDL	128	29.00	84.00	53.20±12.36	47.19±8.70	0.002
LDL	128	58.00	279.00	141.66±36.51	133.20±38.63	0.206
Üre	128	12.00	278.00	37.59±24.74	46.75±39.30	0.117
Kreatinin	128	0.60	13.90	1.11±1.06	1.47±1.62	0.127
Potasyum	128	2.90	5.70	4.46±0.39	4.54±0.48	0.345
HGB	128	9.04	17.50	13.53±1.29	14.88±1.54	<0.001
HCT	128	25.40	50.10	39.29±3.63	42.91±4.42	<0.001
DKB	128	50.00	130.00	88.59±14.67	84.60±13.25	0.110

Veri setindeki çoklu bağlantı probleminin varlığını ortaya koyabilecek birçok yaklaşım vardır. Ancak bunlardan Varyans şişirme faktörü (Variance Inflation Factors-VIF) olarak bilinen VIF değeri,¹³ VIF değerinin tersi olarak da tanımlanabilecek Tolerans değeri¹³ ve en büyük özdeğerin en küçük özdeğere bölümü olarak ifade edilen koşul sayısı değeri,¹⁵ çoklu bağlantının varlığını ve şiddetini ortaya koyabilmektedir. 10'dan büyük VIF değeri, sıfıra yakın tolerans değeri, 30'dan büyük koşul sayısı değeri güçlü çoklu bağlantı probleminin varlığını göstermektedir. Bu bulgulara ilişkin VIF ve Tolerans değerleri Tablo 2'de, özdeğerler ve koşul sayısına ilişkin değerler Tablo 3'de özetlenmektedir.

Tablo 2. VIF ve Tolerans değerleri

Açıklayıcı değişkenler	VIF	Tolerans
Yaş	1.205	0.829
Boy	1.354	0.738
Karın çevresi	1.118	0.894
Glikoz	1.099	0.909
HDL	1.198	0.834
LDL	1.114	0.897
Üre	4.936	0.202
Kreatinin	4.711	0.212
Potasyum	1.208	0.827
HGB	24.442	0.041
HCT	24.837	0.042
Diastolik kan basıncı	1.065	0.9384

Tablo 3. Özdeğerler ve koşul sayıları

No	Özdeğerler	Koşul sayısı
1	2.841	1
2	1.718	1.65
3	1.198	2.37
4	1.161	2.45
5	1.071	2.65
6	0.996	2.85
7	0.893	3.18
8	0.789	3.6
9	0.665	4.27
10	0.535	5.3
11	0.108	26.12
12	0.0205	138.59

Tablo 2'deki HGB ve HCT değişkenlerine ait VIF değerleri 10'dan büyük, tolerans değerleri sıfıra yakın olduğu için ve Tablo 3'deki değerlerden 12 no'lu özdeğere ait koşul sayısı 30'dan büyük çıktığı için veri setinde güçlü çoklu bağlantı problemi olduğu tespit edilmiştir.¹⁵

Bir veri setinde çoklu bağlantı probleminin varlığı LS tahminlerini işaret ve büyüklük olarak etkilemektedir. Bu nedenle veri setine LS, RR ve PCR analizleri uygulanıp sonuçların karşılaştırmalı olarak incelenmesine karar verilmiştir.

Pamukçu ve ark.

LS, RR ve PCR'nin sonuçlarının incelenmesi

LS, RR ve PCR yöntemleri sonucunda elde edilen parametre tahminleri, standart hataları ve parametrelerin anlamlılık testi sonuçları Tablo 4'de verilmiştir. İncelenen üç yöneme ait belirlilik katsayıları (R²)

sırasıyla 0.6003, 0.5876 ve 0.5994 ve hata kareler ortalamaları (HKO) sırasıyla 13.305, 13.515 ve 13.319 olarak elde edilmiştir.

Regresyon katsayılarına t-testi uygulandıktan sonra elde edilen sonuçlar Tablo 4'de verilmiştir.

Tablo 4. LS, RR ve PCR'nin sonuçları

Değişkenler	LS			RR			PCR		
	β_i	SE	P	β_i	SE	P	β_i	SE	p
Sabit	16.834	31.426	0.593	20.300			15.886		
Boy	-6.590	13.589	0.629	-7.053	13.324	0.291	-7.264	13.535	0.721
DKB	1.042	0.087	<0.001	1.023	0.086	<0.001	1.046	0.086	<0.001
Glikoz	0.002	0.028	0.936	0.004	0.027	0.442	0.004	0.027	0.458
HCT	0.749	1.331	0.575	0.425	0.698	0.268	0.094	0.160	0.295
HDL	-0.099	0.117	0.400	-0.094	0.114	0.189	-0.088	0.115	0.792
HGB	-1.886	3.720	0.613	-0.962	1.952	0.307	-0.058	0.462	0.570
Karın çevresi	29.736	11.445	0.011	29.598	11.300	0.005	30.236	11.412	0.005
Kreatinin	3.013	1.855	0.107	2.654	1.613	0.049	3.019	1.856	0.050
LDL	0.021	0.033	0.536	0.019	0.033	0.283	0.019	0.033	0.301
Potasyum	0.382	2.963	0.898	0.231	2.873	0.468	0.614	2.929	0.436
Ure	-0.060	0.079	0.452	-0.048	0.069	0.242	-0.064	0.079	0.802
Yaş	0.146	0.119	0.221	0.139	0.114	0.114	0.158	0.116	0.090
	R ² =0.6003 HKO=13.341			R ² =0.5876 HKO=13.515			R ² =0.5994 HKO=13.305		

Buna göre LS modelinde sadece DKB ve karın çevresi değişkenleri anlamlı iken, yanlış tahmin tekniklerinde çoklu bağlantı probleminin giderilmesi sonucunda kreatinin değişkeni de anlamlı bulunmuştur.

Tablo 4'deki sonuçlara göre göre; LS modeli; Y=SKB, X₁= Karın çevresi, X₂=DKB olmak üzere; Y= 29,736. X₁+1,042.X₂ olarak elde edilmiştir.

Fakat verideki çoklu bağlantı probleminde dolayı karın çevresi katsayısına işaret ve büyüklük olarak

güvenilmemelidir. RR modeli; X₃: Kreatinin olmak üzere Y=29,598.X₁+1,023.X₂+2,654.X₃ olarak elde edilmiştir. Son olarak PCR modeli; Y=30,236.X₁+1,046.X₂+3,019.X₃ olarak elde edilmiştir.

Bu sonuç; RR modeliyle benzerlik göstermektedir. Ridge regresyon yönteminde modelin performansını k ridge parametresi belirlemektedir. k 'nın artan değerleri için katsayılarla ilişkin k , R², HKO ve en büyük VIF değerleri Tablo 5'de verilmektedir.

Tablo 5. Ridge Regresyon için analiz sonuçları

k	R ²	HKO	En büyük VIF
0	0.6	13.34	17.51
0.01	0.59	13.41	11.41
0.02	0.59	13.51	6.61
0.03	0.58	13.61	4.37
0.04	0.58	13.7	3.14
0.05	0.57	13.79	2.46
0.06	0.56	13.88	2.21
0.07	0.56	13.97	2
0.08	0.55	14.05	1.82
0.09	0.55	14.13	1.67
0.1	0.54	14.21	1.53
0.2	0.5	14.89	0.82
0.3	0.46	15.44	0.63
0.4	0.43	15.9	0.52
0.5	0.4	16.28	0.45
0.6	0.38	16.61	0.39
0.7	0.36	16.89	0.34
0.8	0.34	17.14	0.31
0.9	0.32	17.36	0.27
1	3	17.56	0.25

RR modelini göz önüne aldığımızda;

$$\beta^* = (X'X + kI)^{-1}X'Y$$

Sistolik Kan Basıncının Tahmininde Yanlı Regresyon Yöntemlerinin Kullanılması

ifadesinden de anlaşılacağı üzere $k=0$ için RR tahmin edicisi LS'ye eşit olmaktadır. k 'ya 0 ve 1 arasında değerler verilerek en iyi tahmin elde edilmeye çalışılır.

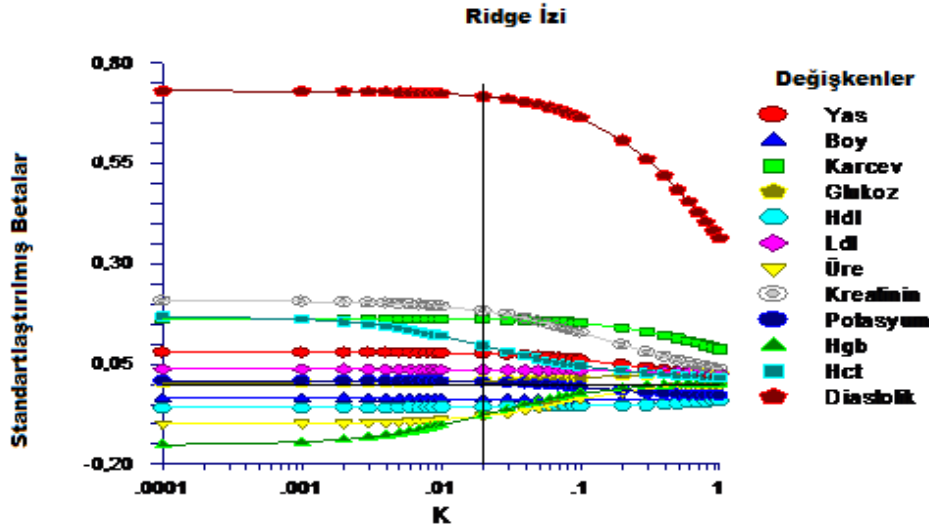
Uygun k değerinde çoklu bağlantı problemi ve katsayılardaki kararsızlık giderilmiş olmalı, fakat R^2 değeri LS'de elde edilen R^2 değerinden fazla uzaklaşmamalıdır.

Bu nedenle katsayıların durağanlaşmaya başladığı, kuramsal beklentilere işaret ve büyüklük olarak uygun

olduğu ve çoklu bağlantının giderildiği ilk değer seçilmelidir. Burada en büyük yardımcı, VIF değeridir.

Her bir açıklayıcı değişken için hesaplanan VIF değerleri 10'dan küçük olmalıdır. Tablo 5'den de görüldüğü gibi $k=0,02$ için en büyük VIF değeri 6,61 olduğu için verideki çoklu bağlantı problemi ortadan kalkmaktadır. Bu nedenle RR analizinde $k=0,02$ seçilmiştir.

Katsayıların durağanlaşmaya başladığının bir ölçütü olarak da Şekil 1'de Ridge İzi verilmiştir.



Şekil 1. Ridge İzi

PCR analizinde ise önemli temel bileşen sayısına karar verebilmek için Tablo 6'daki bilgilerden faydalanılmıştır.

Tablo 6. Temel Bileşenler Regresyonu için analiz sonuçları

PC	R ²	HKO	Max VIF
1	0.0059	20.9821	0.0898
2	0.0181	20.8527	0.2147
3	0.2536	18.1811	0.2311
4	0.3178	17.3816	0.4488
5	0.3752	16.6348	0.4896
6	0.3816	16.549	0.7658
7	0.5262	14.4853	0.8458
8	0.5265	14.4805	0.8507
9	0.5893	13.4866	1.0876
10	0.5942	13.4065	1.3411
11	0.5994	13.3192	4.8951
12	0.6003	13.305	24.8367

Yine en büyük VIF değeri çoklu bağlantının giderildiğinin bir ölçütü olarak ele alınırsa 11. temel bileşen için en büyük VIF değeri 4.89 olduğu için önemli temel bileşen sayısı 11 olarak alınmıştır.

Tartışma

Bir olayın ortaya çıkmasında; etkisi bilinen, etkisi olduğundan şüphelenilen ve etkisinin olup olmadığı henüz bilinmeyen çok sayıda etken (faktör, belirleyici değişkenler) bulunabilir. Bu gibi durumlarda sebep-sonuç ilişkilerini bilimsel olarak ortaya çıkarabilmek için çeşitli istatistik yöntemlerden yararlanmak gerekir. Özellikle sebep-sonuç ilişkisi bilinen durumlarda bile etkenlerin bazılarının majör etken, bazılarının ise minör etken olarak tanımlanması, problemlerin çözülmesinde önceliklerin ortaya konması bakımından büyük önem taşır.

Tıpta klinik, patolojik, radyolojik, laboratuvar vb. bulguları ile bir fenomenin çok sayıda değişkeni ele

Pamukçu ve ark.

alınarak veri toplanmaktadır. Bu değişkenlerin birçoğu birbirleri ile yüksek derecede ilişkili olmakta, bazıları ise fenomen ile çelişir durumda bilgi vermektedirler.

Bu durum tanı ve tedavide tereddütlere, yanılığara yol açabilmektedir. Hangi değişkenlerin problemi incelemede etkin, yardımcı rol oynadıklarını ve hangilerinin bu problemi incelemede etkin rolünün bulunmadığını belirlemede ileri istatistik yöntemlerden yararlanmak faydalı sonuçlar vermektedir.¹⁶

Özellikle yaş ilerledikçe morbidite ve mortalite nedenleri olan kalp yetersizliği, inme, böbrek yetersizliği, koroner kalp hastalığı ve periferik damar hastalığına yol açan hipertansiyonu etkileyen risk faktörlerini belirleyici ve hipertansiyon hastalarını sağlam bireylerden ayırıcı istatistiksel modellerin geliştirilmesine yönelik çalışmalar önem kazanmaktadır.¹

İstatistiksel modelleme-regresyon yöntemleri, sistolik kan basıncı gibi bir değişkeni etkileyebilecek birçok değişkenin bütününe dikkate alarak, istatistik ve klinik olarak önemli risk faktörlerini belirleyebilmektedir.

Açıklayıcı değişkenler birbiriyle bağıntılı ise, bunlardan bazılarının modelden çıkartılması gerekebilir. Ancak hangi değişkenlerin modelden çıkarılacağı özellikle klinik çalışmalarda oldukça önemlidir. Modelden yanlışlıkla bir değişkenin çıkartılması, modelin hatalı tanımlanmasına neden olabilir. Bu kapsamda, birbirleri ile yüksek derecede ilişkili olan değişkenler arasındaki çoklu bağlantı sorununu ortadan kaldırabilen değişik yöntemlerin kullanılması daha etkin sonuçlar verecektir. Bu çalışmada, incelenen açıklayıcı değişkenler arasında güçlü çoklu bağlantı probleminin var olduğu istatistik yöntemler (koşul sayısı>30 gibi) ile belirlenmiştir. Bu nedenle sistolik kan basıncının tahmini için bazı biyokimyasal değişkenler ile kardiyovasküler risk faktörleri arasında yer alan kimi faktörlerin bulunduğu regresyon modeline, LS, RR ve PCR yöntemleri uygulanmış ve sonuçlar değerlendirilmiştir.

Hipertansiyon erken tanı ile kontrol altına alınabilen bir hastalık olduğu için değişik risk faktörlerinin belirlenmesi ile gelişimi ve ilerlemesine karşı önlemler alınabilir. Türkiye’de hipertansiyonun tahmini ve ilişkili faktörlerin belirlenmesine yönelik olarak ileri istatistik yöntemleri ele alan ve karşılaştıran çok fazla çalışma bulunmamaktadır. Türe ve arkadaşlarının “Hipertansiyonun tahmini için çoklu tahmin modellerinin karşılaştırılması” isimli çalışmalarında, primer hipertansiyonun tahminlenmesinde lojistik regresyon analizi (LR), esnek diskriminant analizi (EAA), ve yapay sinir ağları (YSA), yöntemlerini karşılaştırmışlardır.

Sonuç olarak, aile hikâyesi, lipoprotein A, trigliserid, sigara kullanımı ve vücut kitle indeksi değişkenlerinin, kontrol ve hipertansiyonlu hasta gruplarını tahmin etmede kullanılabileceğine ve YSA’nın performansının LR ile EAA’dan istatistiksel olarak daha iyi olduğuna karar vermişlerdir.¹ Bu çalışmada, RR ve PCR analizleri sonucunda, sistolik kan basıncını tahmin etmek için istatistiksel olarak önemli faktörler; kanın çevresi, kreatinin ve diastolik kan basıncı olarak bulunmuştur. Benzer olarak, Bajraktari et al. (2009) kreatinin düzeyinin hipertansiyon kontrolü için bağımsız prediktörlerden biri olduğunu bildirmiştir.¹⁷

Bu sonuç, var olan önemli değişkenlere ek olarak hipertansiyonun kontrol altına alınabilmesi açısından dikkate alınmalıdır. Modelde yer alan önemli değişkenler, bireyleri daha hasta olmadan korumak, aydınlatmak ve koruyucu tedbirler alınabilmeleri açısından önemli olabilmektedir.

Sonuç ve Öneriler

Çoklu bağlantı problemi, LS tahminlerini işaret ve büyüklük olarak etkilemektedir. Çoklu bağlantı sonucunda artan varyans, LS tahminlerinin güven aralıklarını genişletmekte ve bu da farklı örnekler için farklı tahminlerin elde edilmesine sebep olmaktadır. Ayrıca F testi sonucunda model tümel olarak anlamlı iken, t-testi sonucunda tahminlerin büyük bir çoğunluğu anlamsız çıkabilmektedir.

Sonuç olarak; incelenen problemlerde çoklu bağlantılı açıklayıcı değişkenlerin modelden çıkartılması istenmiyor ve çoklu bağlantı probleminin giderilmesi de gerekli ise, RR ve PCR gibi yanlı tahmin yöntemlerinin kullanılması tavsiye edilebilir.

Bu çalışmada çoklu bağlantılı verilere her üç yöntem uygulanmış ve sonuçta PCR ve RR yöntemleri ile daha tutarlı ve durağan kuramsal beklentilere uygun tahminler elde edildiği belirlenmiştir. Modellerin birbirlerine üstünlüklerini belirlemede ölçü olarak R² belirlilik katsayısı ve HKO kullanılmıştır. Bu nedenle PCR ve RR sonuçları incelendiğinde, PCR’nin RR’ye göre daha iyi R² ve daha düşük HKO’na sahip tahminler sağladığı belirlenmiştir.

Kaynaklar

1. Türe M, Kurt İ, Yavuz E, Kürüm T. [Comparison of multiple prediction models for hypertension (Neural networks, logistic regression and flexible discriminant analyses)] *Anadolu Kardiyoloji Dergisi* 2005;5(1):24-8.
2. Müller DN, Luft CF, Direct renin inhibition with aliskiren in hypertension and target organ damage. *Clin J Am Soc Nephrol* 2006;1:221-8.
3. Çakmak HA, Arslan E, Erdine S, [Unmet needs in hypertension], *Türk Kardiyoloji Derneği Araştırması. Arch Turk Soc Cardiol*, 2009;37(7):1-4.

Sistolik Kan Basıncının Tahmininde Yanlı Regresyon Yöntemlerinin Kullanılması

4. El-Guindy MS, Khatilo MNO, Clinical guidelines for the management of hypertension. 2005. EMRO technical publications series 29. WHO Regional Office for the Eastern Mediterranean Cairo.
5. Alpar R, Çoklu Doğrusal Regresyon, Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş. 12. Baskı, Ankara:Nobel Yayın Dağıtım 2003;299-301.
6. Karagöz M, Regresyon Analizi, İstatistik Yöntemleri. 6. Baskı, Bursa: Ekin Kitabevi 2006;313-6.
7. Freund JE. Regresyon ve Korelasyon (Bağlanım ve İlişki), Matematiksel İstatistik Çeviren: Ümit Şenesen, İstanbul: Literatür Yayıncılık 2002;456-507.
8. Orhunbilge N. Çoklu Regresyon ve Korelasyon Analizi. Uygulamalı Regresyon Ve Korelasyon Analizi. 2. Baskı, İstanbul: İ.Ü. Basım ve Yayınevi Müdürlüğü, 2002;99-100.
9. Albayrak AS. [Biased estimation techniques which are the alternative of least square technique for multicollinearity problem and an application], ZKÜ Sosyal Bilimler Dergisi 2005;1(1):105-26.
10. Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. Technometrics 1970a; 12(1):55-67.
11. Hoerl AE, Kennard RW. Ridge Regression: Applications to Non-Orthogonal Problems. Technometrics 1970b; 12 (1):69-82.
12. Özkan K. [Field water capacity modelling according to soil Texture using principle component regression analysis.], SDÜ Orman Fakültesi Dergisi. 2009;2(2):1-9.
13. Tatlıdil H. Temel Bileşenler Analizi, Uygulamalı Çok Değişkenli İstatistiksel Analiz 1. Baskı, Ankara Ziraat Matbaacılık AŞ 2002;138-62.
14. Sümbüloğlu K. Sümbüloğlu V. Güney Z. Klinik ve Saha Araştırmalarında Örneklem Yöntemleri ve Örneklem Büyüklüğü, Ankara: Alp Ofset Matbaacılık 2005.
15. Gujarati ND. Klasik Modelin Varsayımlarının Genişletilmesi, Temel Ekonometri Çevirenler: Ümit Şenesen. Gülay Göktürk Şenesen, 6.baskı, İstanbul: Literatür Yayıncılık, 2009;338.
16. Özdamar K. [Hypertension and prediction models.] Anadolu Kardiyoloji Dergisi, 2005;5(1):29.
17. Bajraktari G, Sylejmani X, Thaçi K, Elezi S, Ndrepepa G. The quality of arterial hypertension treatment in cardiology service in Kosova-a single center study. Anadolu Kardiyoloji Dergisi 2005;9:96-101.

İletişim Adresi: Dr. Esra PAMUKÇU

Fırat Üniversitesi Fen Fakültesi,
İstatistik Bölümü, ELAZIG
Tel: 0424 2370000-3576
Faks: 0 424 2330062
e-mail: esra_pamukcu@hotmail.com,
e-mail: epamukcu@firat.edu.tr