

## Lojistik Regresyon Analizi Yardımıyla Denekte Menopoz Evresine Geçiş İlişkin Bir Sınıflandırma Modelinin Elde Edilmesi\*

Neslihan İYİT<sup>1</sup> , Aşır GENÇ<sup>1</sup>

**Özet:** Menopoz; kadında yumurtalıkların östrojen yapımını kesmelerinden itibaren gelişen ve kadının doğurganlık döneminden kısırılık dönemine geçişini tanımlayan bir evredir. Yapılan bu çalışmada, temel olarak menopozun varlığı veya yokluğunun belirlenebilmesi için kurulacak lojistik model yardımıyla, deneklerin bu fiziksel durumdan hangisine sahip olduklarını tahmin etmeye yardımcı olan bir sınıflandırma modeli elde etmek hedeflenmiştir. Yaş, E2 hormon seviyesi, kürtaj sayısı, operasyon özgeçmiş, ateş-ter basması, uykusuzluk problemi, idrar sorunu ve kanama problemi değişkenleri denekte menopoz sorunu ile karşılaşma durumunu etkilediği düşünülen olası risk faktörleri olarak ele alınmış ve kurulması öngörülen lojit model yardımıyla anlamlı bulunmuştur. Uygulanan lojistik regresyon modeli; menopoz şikayetleri ile Konya bölgesinin kırsal ve kent kesiminden özel bir kliniğe başvuran 115 hastaya ilişkin verileri %93 oranında başarıyla sınıflandırmıştır.

**Anahtar Kelimeler:** Menopoz, lojistik regresyon analizi, Wald test istatistiği, Box-Tidwell yaklaşımı, Hosmer-Lemeshow  $\hat{C}_g$  istatistiği.

## Obtaining a Classification Model Relating to Transition Stage to Menopause In an Individual by Using Logistic Regression Model

**Abstract:** Menopause is a stage in any female's life period defined as transition from prolificacy period to sterility (bareness) period develops with stopping ovaries producing eustrogen hormone. In this study ,basically, it is aimed to obtain a classification model for the purpose of estimating which physical situation does an individual have in determining presence or absence of menopause by a logistic model. In the model building stage; risk factors such as age, E2 hormone level, number of curettage, gynecologic operation history, hot feeling-sweating, sleeping problem, uriner problems and hemorrhage problem are found significant to determine an individual having menopause problem by a logit model. As a result of this study, it is concluded that the logistic regression model classified 115 patients data ,coming to a private clinique from both rural and urban area of Konya with menopause complaints, as in menopause stage or not successively by the ratio 93%.

**Key Words:** Menopause, logistic regression analysis, Wald test statistics, Box-Tidwell Approximation, Hosmer-Lemeshow's  $\hat{C}_g$  statistics.

---

\* Bu makale Yüksek Lisans Tezinin bir parçasıdır.

<sup>1</sup> Selçuk Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 42031 Kampüs/KONYA

## Giriş

Menopoz; kadında yumurtalıkların östrojen yapımını kesmelerinden itibaren gelişen (sıklıkla 45–55 yaşları arasında) ve hayat boyu devam eden bir evre olarak tanımlanır. Menopoz bazen “Hayat Değişimi” olarak da ifade edilir ve kadının doğurganlık döneminden kısırılık dönemine geçişini tanımlar. Menopoz döneminde kadınların çoğu sıcak basmaları, gece terlemeleri, uyku bozuklukları, çarpıntı, eklem ve kas ağrıları ile birlikte huzursuzluk şikayetleri, depresyon, sıkıntı hissi, hafızanın zayıflaması ve değişen ruh hali gibi psikolojik semptomlar olmak üzere değişen derecelerde menopozal belirtiler yaşarlar. Öte yandan hormonal değişikliklerle de ilişkili olan üriner değişiklikler de menopozun diğer semptomlarına eşlik edebilir.

Dolayısıyla kadındaki üretkenlik evresinin bitimiyle beraberinde osteoporoz gibi birçok sorunu da getiren menopoza girmede önlenemez risk faktörlerinin belirlenmesi, menopoz yaşını ileri çekebilme için yöntemlerin geliştirilmesine katkıda bulunması açısından önemlidir.

## Metodoloji

Bağımlı değişkenin sürekli olduğu ve başta artıklar üzerinde normallik olmak üzere çeşitli varsayımların yapıldığı regresyon modelinde, bağımlı değişkenin iki yada ikiden çok düzey içeren kesikli bir değişken olması durumunda lineer regresyon modeli için öngörülen varsayımlar bozulmakta ve hata teriminin binom dağılım göstermesi sebebiyle gözlem varyansları eşit olmamaktadır. Bu durumda veriler lineer regresyon analizi ile incelenemediğinden lojistik regresyon analizinin kullanılması önerilmektedir [2].

Lineer regresyon için koşullu ortalamanın;

$$E(Y/x) = \beta_0 + \beta_1 x \quad (1)$$

eşitliği yardımı ile verilen ifadesinden bağımsız değişkenin  $-\infty$  ile  $+\infty$  arasında değişen sürekli bir değişken olmasından dolayı  $E(Y/x)$ 'nin mümkün olan her değeri alabileceği görülür. Ancak bağımlı değişkenin iki düzeyli olduğu durumda koşullu ortalama sıfır ile bir arasında değişmek zorundadır  $[0 \leq E(Y/x) \leq 1]$  [3]. Lojistik regresyon analizinde ise bağımsız değişkenin/değişkenlerin değeri bilindiğinde  $Y$ 'nin koşullu ortalaması  $\pi(x) = E(Y/x)$  ifadesi şeklinde olup, (1) eşitliğin sol tarafı sıfır ile bir arasında sınırlı olasılık değerleri almasına rağmen bağımsız değişkenlerin sonsuz sayıda değer alabilmesinden dolayı söz konusu eşitlik her zaman sağlanamamaktadır. Böylesi bir durumu ortadan kaldırmak için en iyi çözüm bağımlı değişken olarak ifade edilen olasılık değeri olan  $\pi(x)$ 'i  $-\infty$  ile  $+\infty$  arasında tanımlı hale getirebilecek bir transformasyonla mümkündür. Bu durumda kullanılacak lojistik regresyon modelinin açık şekli ise;

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2)$$

şeklinde dir.

Eşitlik (2) ile verilen olasılık modelinde olasılık değerleri üzerinde yapılacak  $\pi(x)/[1-\pi(x)]$  dönüşümü bağımlı değişkenin sınırlarını  $(0, \infty)$  arasında yapacaktır. Sınırları  $(-\infty, +\infty)$  arasında yapmak içinse bu oranın doğal logaritması alınır. Böylece elde edilen yeni bağımlı değişkeni bağımsız değişkenin lineer bir fonksiyonu olarak (3) eşitliği ile gösterildiği gibi yazabilmek mümkün olacaktır [4]. Bu şekilde  $\pi(x)$ 'i  $-\infty$  ile  $+\infty$  arasında tanımlı hale getiren bu transformasyona lojit transformasyonu denir. Lojit transformasyonu  $\pi(x)$  cinsinden;

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] \quad (3)$$
$$= \beta_0 + \beta_1 x$$

şeklinde dir. Bu transformasyon yardımıyla elde edilen modele lojit adı verilir ve  $g(x)$  ile gösterilir. Lojit olarak adlandırılan  $g(x)$  fonksiyonu parametreleri bakımından lineer olup bağımsız değişkenin aldığı değerlere bağlı olarak  $-\infty$  ile  $+\infty$  arasında değişen sürekli bir fonksiyon olarak elde edilir [5].

Görüldüğü üzere lineer regresyon analizinde (1) eşitliği yardımı ile verilen bağımlı değişkenin koşullu beklenen değeri/ortalaması hesaplanırken, lojistik regresyon analizinde ise (2) eşitliği yardımı ile verilen lojistik regresyon modeli kullanılarak bağımlı değişkenin alacağı değerlerden birinin gerçekleşme olasılığı hesaplanır. [6].

Lojistik modelin biyolojik deneyler (bioassay) analizi için kullanımı ilk olarak [7] tarafından önerilmiş, [8] bu modeli gözden geçirerek çeşitli uygulamalarını yapmış, özet gelişmeler ise [9] ve [10] tarafından verilmiştir. Ayrıca verilerin lojistik modele uyumu ile ilgili bir çok çalışma literatürde yer almaktadır. Bunlar arasında [11], [12] ve [13] tarafından yapılanlar en önemlileridir. [14] iki grup lojistik modellerde etkin (influential), aykırı (outlier) gözlemleri ve teşhis ölçütlerini (diagnostics), [15] ve [16] ise çoklu grup lojistik modellerde etkin ve aykırı gözlemlerle teşhis ölçütlerini incelemiştir.

Bu çalışmada uygulamaya dahil edilen 1999 ile 2003 yılları arasında menopoza şikayeti ile özel bir kliniğe başvuran 115 denek yapılan tetkik sonuçlarına göre menopoza var ya da yok olarak sınıflandırılmıştır [1]. Yapılan bu çalışmada menopoza ile ilişkili olabileceği düşünülen risk faktörleri çok değişkenli lojistik regresyon analiziyle belirlenmeye çalışılacaktır.

### **Menopoza Verilerine İlişkin Çok Değişkenli Lojistik Modelin Elde Edilmesi**

Menopoza girmede önlenilebilir risk faktörlerinin belirlenmesi, menopoza yaşını ileri çekebilmek için yöntemlerin geliştirilmesine katkıda bulunması açısından önemlidir. Dolayısıyla menopoza etki eden risk faktörlerinin belirlenmesi amacıyla yönelik lojistik regresyon analizi kullanılarak yapılan bu çalışmada uygulamaya dahil edilen 115 denek üzerinde yapılan menopoza tetkikleri sonuçlarına göre menopoza var ya da yok olarak sınıflandırılmıştır [1]. Yapılan bu sınıflandırma, lojistik regresyon analizinde bağımlı değişkeni belirlemek için kullanılmış olup bağımlı değişken  $Y$ ; hastada menopoza evresinin varlığını ( $Y_i = 1$ ) veya yokluğunu ( $Y_i = 0$ ) gösteren ve iki değer alabilen kategorik (kesikli, sınıflı) bir değişkendir. Çalışmada lojistik modele girmeye aday değişkenler olarak belirlenen hastanın corpus, endometrium ve uterus muayeneleri, enfeksiyon varlığı, ateş basması-terleme şikayeti, yaşam şekli, psikolojik problemleri, uykusuzluk sorunu, idrar yolları problemleri, ağrı şikayeti, kanama sorunu ve operasyon özgeçmişini kategorik yapıdaki bağımsız değişkenler (çalışma faktörleri); hastanın yaşı, son doğum yaşı, açlık kan şekeri (AKŞ), trombosit, E2 hormon seviyesi değerleri sürekli ölçekli değişkenler; hastanın doğum sayısı, düşük sayısı, kürtaj sayısı ise kesikli ölçekli değişkenler olarak alınmıştır.

Ele alınan verilere ilişkin çok değişkenli lojistik regresyon modelinin kurulması aşamasında ilk adım olarak modele girmeye aday değişkenlerin belirlenebilmesi amacıyla her bir değişkenin tek değişkenli lojistik regresyon analizi yardımıyla katsayılarının anlamlılık testinin yapılması gerekir. Bağımlı değişken olan menopoza varlığı ile ilişkili olabileceği düşünülen olası değişkenlerin tek değişkenli lojistik regresyon analizi sonuçları Tablo 1’de verilmiştir.

Tablo 1. Menopozun varlığı ile ilişkili olabileceği düşünülen değişkenlerin tek değişkenli lojistik regresyon modellerine ilişkin sonuçları

Değişken	$\hat{\beta}$	$SE(\hat{\beta})$	Wald	sd	Anlamlılık	$\hat{\psi}$	%95Güven Aralığı	-2(Log-olabilirlik)	Doğru sınıfl. oranı
YAŞ	0.153	0.042	13.385	1	0.000*	1.166	(1.074, 1.265)	125.832	0.713
SONDOĞ	0.049	0.037	1.698	1	0.193*	1.050	(0.976, 1.130)	137.362	0.673
AKŞ	0.021	0.011	3.742	1	0.053*	1.022	(1.000, 1.044)	137.326	0.687
TROMBOSİT	0.000	0.000	0.612	1	0.434**	1.000	(1.000, 1.000)	142.339	0.687
E2	-0.061	0.015	17.438	1	0.000*	0.941	(0.915, 0.968)	101.847	0.791
DOĞUMSAYISI	0.161	0.108	2.252	1	0.133*	1.175	(0.952, 1.451)	140.537	0.687
DÜŞÜK SAYISI	-0.143	0.233	0.376	1	0.540**	0.867	(0.549, 1.368)	142.581	0.687
KÜRTAJSAYISI	-0.201	0.129	2.453	1	0.117*	0.818	(0.635, 1.052)	140.493	0.678
OPERE(1)	-0.931	0.481	3.749	1	0.053*	0.394	(0.154, 1.012)	138.838	0.687
YAŞAM ŞEKLİ(1)	-0.325	0.410	0.631	1	0.427**	0.722	(0.324, 1.612)	142.312	0.687
ATEŞTER(1)	2.872	0.510	31.774	1	0.000*	17.676	(6.511,47.985)	102.185	0.809
PSİKO(1)	1.031	0.435	5.624	1	0.018*	2.805	(1.196, 6.579)	136.941	0.687
UYKUSUZLUK(1)	3.062	0.762	16.146	1	0.000*	21.371	(4.799, 95.166)	112.094	0.687
IDRAR(1)	0.972	0.446	4.746	1	0.029*	2.643	(1.102, 6.335)	137.844	0.687
AĞRI(1)	0.432	0.411	1.107	1	0.293**	1.541	(0.689, 3.447)	141.846	0.687
KANAMA(1)	-1.402	0.452	9.602	1	0.002*	0.246	(0.101, 0.597)	133.176	0.713
UTERUS	-	-	1.703	2	0.427**	-	-	141.285	0.687
UTERUS(1)	-0.470	1.313	0.128	1	0.720**	0.625	(0.048, 8.201)	141.285	0.687
UTERUS(2)	0.216	1.246	0.030	1	0.863**	1.241	(0.108, 14.259)	141.285	0.687
ENDOMETR	-	-	5.231	2	0.073*	-	-	137.553	0.687
ENDOMETR(1)	-1.099	1.219	0.812	1	0.368**	0.333	(0.031, 3.638)	137.553	0.687
ENDOMETR(2)	-0.047	1.181	0.002	1	0.969**	0.955	(0.094, 9.661)	137.553	0.687
CORPUS	-	-	4.285	3	0.232*	-	-	121.333	0.713
CORPUS(1)	-2.303	1.151	4.001	1	0.045*	0.100	(0.010, 0.955)	121.333	0.713
CORPUS(2)	-0.862	0.825	1.092	1	0.296**	0.422	(0.084, 2.126)	121.333	0.713
CORPUS(3)	7.816	22.292	0.123	1	0.726**	2481.06	(0.000, 2.E+22)	121.333	0.713
ENFEKTE(1)	0.836	0.412	4.129	1	0.042*	2.308	(1.030, 5.171)	138.716	0.687

Tablo 1'de yer alan (\*) simgesi tek değişkenli lojistik regresyon analizi sonucunda istatistiksel olarak önemli değişkenleri göstermek amacıyla, (\*\*) simgesi ise istatistiksel olarak önemli olmayan değişkenleri göstermek amacıyla kullanılmıştır.

Tablo 1.'den yararlanarak çok değişkenli lojistik regresyon modeline girmeye aday değişkenlerin seçiminde her bir modelde yer alan değişkenlere ait katsayıların ayrı ayrı önem kontrolünün yapılabilmesi için Wald test istatistiğinden yararlanır. Bu amaç için kullanılacak hipotezler ise şöyledir;

$$\begin{aligned} H_0 &= \beta_i = 0 \\ H_1 &= \beta_i \neq 0 \end{aligned} \quad (4)$$

Lojistik regresyon analizinde tek değişkenli modeller yardımıyla istatistiksel olarak önemli değişkenlerin seçiminde kullanılan Wald test istatistiğinin olasılık değeri 0.25 olarak belirlenen yanılma düzeyinden küçük çıktığı takdirde ( $p < 0.25$ ) ilgili değişkenlerin çok değişkenli modele aday değişkenler olarak belirlenebileceği [17] ve [18] tarafından yapılan çalışmalarda ortaya konulmuştur. Böylece Tablo 1 yardımıyla her bir tek değişkenli modelde yer alan değişkenlerin katsayılarının anlamlılık testinin yapılması sonucunda trombosit, düşük sayısı, yaşam şekli, ağrı, uterus, endometrium, corpus değişkenleri dışındaki tüm değişkenler bağımlı değişkenle istatistiksel olarak anlamlı bir ilişki içinde olduklarından dolayı çok değişkenli modele aday değişkenler olarak belirlenmişlerdir. Aday değişkenlerle kurulan çok değişkenli modele ilişkin sonuçlar Tablo 2'de verilmiştir.

**Tablo 2. Tek değişkenli modelde önemli olarak tanımlanan aday değişkenleri kapsayan çok değişkenli lojistik regresyon modeli sonuçları**

Değişkenler	$\hat{\beta}$	$SE(\hat{\beta})$	Wald	sd	Anlam.	$Exp(\hat{\beta})$	$Exp(\hat{\beta})$ için 95,0% güven aralığı	
							Alt	Üst
YAŞ	,207	,113	3,376	1	,066	1,230	,986	1,534
SONDOĞUM	,021	,098	,044	1	,834	1,021	,842	1,238
AKŞ	,020	,030	,451	1	,502	1,021	,962	1,083
E2	-,071	,025	8,172	1	,004	,931	,887	,978
DOĞSA	-,121	,291	,172	1	,678	,886	,501	1,567
KÜRTAJSA	,328	,318	1,067	1	,302	1,388	,745	2,588
OPERE(1)	-2,356	1,503	2,457	1	,117	,095	,005	1,804
ATESTER(1)	3,786	1,198	9,981	1	,002	44,074	4,209	461,542
PSIKO(1)	,330	1,119	,087	1	,768	1,391	,155	12,474
UYKUSUZL(1)	3,937	1,839	4,586	1	,032	51,287	1,396	1883,969
IDRAR(1)	1,563	1,129	1,916	1	,166	4,773	,522	43,635
KANAMA(1)	-2,345	1,175	3,981	1	,046	,096	,010	,959
ENFEKTE(1)	,428	1,119	,147	1	,702	1,535	,171	13,753
Sabit	-10,186	5,995	2,887	1	,089	,000		

Çok değişkenli lojistik regresyon modelinin kurulması aşamasında ikinci adım olarak modelde yer alması düşünülen aday değişkenlerin seçimi; adimsal seçim metotlarından olabilirlik oran test istatistiğine dayalı geriye doğru eleme yöntemi (backward LR elimination method) kullanılarak yapılmış ve lojistik modele ilişkin elde edilen sonuçlar Tablo 3'te verilmiştir.

**Tablo 3. Adimsal seçim metotlarından olabilirlik oran test istatistiğine dayalı geriye doğru eleme yöntemi (backward LR elimination method) kullanılarak elde edilen çok değişkenli lojistik model sonuçları**

Değişkenler	$\hat{\beta}$	$SE(\hat{\beta})$	Wald	sd	Anlaml.	$Exp(\hat{\beta})$	$Exp(\hat{\beta})$ için 95,0% güven aralığı	
							Alt	Üst
YAŞ	,196	,094	4,365	1	,037	1,216	1,012	1,462
E2	-,075	,026	8,157	1	,004	,928	,881	,977
KÜRTAJSA	,377	,296	1,615	1	,204	1,457	,815	2,605
OPERE(1)	-2,524	1,398	3,260	1	,071	,080	,005	1,241
ATESTER(1)	3,936	1,146	11,796	1	,001	51,199	5,418	483,816
UYKUSUZL(1)	4,161	1,830	5,169	1	,023	64,124	1,775	2316,493
IDRAR(1)	1,848	1,018	3,299	1	,069	6,350	,864	46,666
KANAMA(1)	-2,486	1,117	4,949	1	,026	,083	,009	,744
Sabit	-7,188	4,742	2,298	1	,130	,001		

Çok değişkenli lojistik regresyon modeline girecek değişkenlerin seçimini gerçekleştirme işleminden sonra model kurmada üçüncü adım olarak modele giren sürekli değişkenlerin lojitle lineer bir ilişki içinde olup olmadıklarının ve modele doğru ölçek kullanılarak girip girmediklerinin kontrol edilmesi aşaması gelir [19]. Modelde yer alan yıl cinsinden ölçülmüş sürekli değişken olan yaş değişkeni ile mIU/ml cinsinden ölçülmüş sürekli bir değişken olan E2 hormon seviyesini temsil eden E2 değişkeninin lojitle lineerliği test edilecektir. Tablo 4'ten görüldüğü üzere Box-Tidwell yaklaşımı kullanılarak, sürekli dağılıma sahip olan yaş ve E2 değişkenini kapsayan çok değişkenli lojistik modele YAŞ[ln(YAŞ)] ve E2[ln(E2)] değişkenlerinin eklenmesi sonucunda, bu yeni değişkenlerin istatistiksel olarak önemli bir eğime sahip olması yaş ve E2 sürekli değişkenlerinin lojitle lineer bir ilişki içinde olmadıkları varsayımını desteklemektedir. Bu nedenle yaş ve E2

değişkenlerinin modele sürekli birer değişken olarak girmelerinde sakınca vardır. Bu durumda bu değişkenlerin modele dizayn değişkenleri kullanılarak girmeleri gerekmektedir.

**Tablo 4. Box-Tidwell yaklaşımı kullanılarak sürekli değişkenler için doğru ölçeklerin elde edilmesi amacıyla kurulan çok değişkenli lojistik model sonuçları**

Değişkenler	$\hat{\beta}$	$SE(\hat{\beta})$	Wald	sd	Anlaml.	$Exp(\hat{\beta})$	$Exp(\hat{\beta})$ için 95,0% güven aralığı	
							Alt	Üst
YAŞLN(YAŞ)	,041	,020	4,394	1	,036	1,042	1,003	1,083
E2LN(E2)	-,017	,006	7,634	1	,006	,983	,971	,995
KÜRTAJSA	,390	,303	1,657	1	,198	1,477	,816	2,674
OPERE(1)	-2,559	1,419	3,252	1	,071	,077	,005	1,249
ATESTER(1)	4,064	1,158	12,315	1	,000	58,225	6,016	563,569
UYKUSUZL(1)	4,236	1,912	4,912	1	,027	69,160	1,632	2930,828
IDRAR(1)	1,844	1,037	3,163	1	,075	6,324	,828	48,278
KANAMA(1)	-2,606	1,140	5,231	1	,022	,074	,008	,689
Sabit	-5,890	3,953	2,220	1	,136	,003		

Menopoza girmede önemli bir risk faktörü olarak belirlenen yaş değişkeninin sürekli bir değişken olmasına rağmen "sürekli değişkenler lojitle lineer bir ilişki içinde olmalıdır" şeklindeki lojistik regresyon varsayımını taşımadığı için dizayn değişkenlerinin belirlenerek modele girmesi amacıyla 50 yaş değeri kesim noktası olarak kullanılacaktır. Böylece risksiz grup (referans grup) 50 yaşın altı, riskli grup ise 50 yaş ve üzeri olarak belirlenmiş olacaktır. Yaş değişkeni yerine kullanılacak dizayn değişkeni YAŞ (1) ile gösterilecektir ve bu yeni değişken 50 yaşın altı için 0 ve üzeri için 1 olarak kodlanacaktır. Aynı şekilde E2 değişkeninin içinde dizayn değişkenlerinin belirlenerek modele girmesi amacıyla 40 mlU/ml değeri kesim noktası olarak alınmış böylece risksiz grup (referans grup) 40 mlU/ml değerinin üstü olacak şekilde 0 değeri ile, riskli grup ise 40 mlU/ml değerinin altı olacak şekilde 1 değeri ile kodlanarak oluşturulmuştur. Buna göre yaş ve E2 değişkenlerinin artık kesikli birer değişken olmak üzere kategorilerinin dizayn değişkenleri kullanılarak girdiği çok değişkenli lojistik regresyon modeline ait sonuçlar Tablo 5'te verilmiştir.

**Tablo 5. Sürekli değişkenler için doğru ölçekler elde edildikten sonra kurulan çok değişkenli lojistik model sonuçları**

Değişkenler	$\hat{\beta}$	$SE(\hat{\beta})$	Wald	sd	Anlaml.	$Exp(\hat{\beta})$	$Exp(\hat{\beta})$ için 95,0% güven aralığı	
							Alt	Üst
YAŞ(1)	1,673	,944	3,142	1	,076	5,328	,838	33,874
E2(1)	3,315	,995	11,102	1	,001	27,532	3,916	193,558
KÜRTAJSA	,441	,273	2,622	1	,105	1,555	,911	2,652
OPERE(1)	-2,690	1,268	4,499	1	,034	,068	,006	,815
ATESTER(1)	3,666	1,016	13,008	1	,000	39,094	5,332	286,630
UYKUSUZL(1)	2,888	1,290	5,012	1	,025	17,960	1,433	225,140
IDRAR(1)	1,757	,930	3,568	1	,059	5,796	,936	35,892
KANAMA(1)	-1,764	,956	3,403	1	,065	,171	,026	1,116
Sabit	-3,153	1,437	4,812	1	,028	,043		

Ele alınan çalışma verileri için çok değişkenli lojistik model kurulduktan sonra Tablo 5 ile verilen yeni modeldeki bütün değişkenlerin bağımlı değişkenle istatistiksel olarak anlamlı bir ilişki içinde olduğu görülmektedir.

Çok değişkenli lojistik regresyon modelinin kurulmasındaki bir sonraki adım ise muhtemel etkileşim terimlerinin şekillendirilip modele dahil edilmesi aşamasıdır. Bu çalışmada çok değişkenli lojistik regresyon analizi ile modele herhangi bir katkısının olup olmayacağı incelenen 18 tane muhtemel etkileşim teriminden hiçbiri modele dahil edilmek üzere istatistiksel olarak anlamlı bulunmamış ve etkileşim terimlerinin modelde yer almaması gerektiğine karar verilmiştir.

Yapılan çalışmaya ilişkin Tablo 5 ile verilen çok değişkenli lojistik regresyon analizi sonuçlarına göre denekte menopoza sorunu ile karşılaşma durumuna etki eden risk faktörlerinin modellenmesi aşağıdaki lojit model yardımı ile elde edilmiştir.

$$g(x) = -3.153 + 1.673(YAŞ(1)) + 3.315(E2(1)) + 0.441(KÜRТАJSAY) - 2.690(OPERE(1)) \\ + 3.666(ATESTER(1)) + 2.888(UYKUSUZLUK(1)) + 1.757(İDRAR(1)) - 1.764(KANAMA(1))$$

Ele alınan çalışmada tüm hastalara ilişkin veriler üzerinde yapılan analizlerden elde edilen sonuçlar ışığında menopoza şikayetleri ile hekime gelen bir hastanın menopoza sorunu ile karşılaşma olasılığını belirlemede kullanılacak olasılık değerleri ise aşağıdaki eşitlik yardımıyla;

$$\pi(x) = 1 / \{1 + \exp[-3.153 + 1.673(YAŞ(1)) + 3.315(E2(1)) + 0.441(KÜRТАJSAY) - 2.690(OPERE(1)) \\ + 3.666(ATESTER(1)) + 2.888(UYKUSUZLUK(1)) + 1.757(İDRAR(1)) - 1.764(KANAMA(1))]\}$$

elde edilir.

Kurulan çok değişkenli lojistik regresyon modelinin uyum iyiliğine karar vermek için Tablo 6 ile verilen 20 gözeye ait gözlenen ve beklenen frekanslar kullanılarak Hosmer-Lemeshow  $\hat{C}_g$  istatistiği,  $o_{kl}$  değerleri menopoza girmiş ve girmemiş deneklere ilişkin her bir risk grubu içindeki gözlenen frekansları,  $e_{kl}$  değerleri ise menopoza girmiş ve girmemiş deneklere ilişkin her bir risk grubu içindeki beklenen frekansları göstermek üzere;

$$\hat{C}_g^* = \sum_{k=0}^1 \sum_{l=1}^{10} \frac{(o_{kl} - e_{kl})^2}{e_{kl}} \quad (5)$$

şeklinde hesaplanır [20].

**Tablo 6. Menopoza girmiş olan ve olmayan denekler için her bir onlu risk grubundaki gözlenen ve beklenen frekanslar**

Değer	Gruplar											Toplam
	Frekans	1	2	3	4	5	6	7	8	9	10	
1	Gözlenen	0	1	3	7	11	11	11	13	15	7	79
	Beklenen	0.1	0.7	3.1	8.0	9.8	11.6	10.9	13.0	15.0	7.0	
0	Gözlenen	11	11	8	5	0	1	0	0	0	0	36
	Beklenen	10.9	11.3	7.9	4.0	1.2	0.4	0.1	0.0	0.0	0.0	
Toplam		11	12	11	12	11	12	11	13	15	7	115

Verilen bu bilgiler ışığında Hosmer-Lemeshow  $\hat{C}_g^*$  istatistiğinin değeri 3.041 olarak bulunmuştur. Hosmer-Lemeshow  $\hat{C}_g^*$  istatistiği,  $\alpha = 0.05$  yanılma (önem) düzeyinde ve 8 serbestlik derecesinde ki-kare dağılımıyla karşılaştırılınca, Hosmer-Lemeshow  $\hat{C}_g^*$  istatistiğinin değerinin daha küçük olduğu görülür ( $\hat{C}_g^* = 3.041 < \chi_{0.05}^2(8) = 7.344, p > 0.05$ ). Bu sonuç kurulan modelin uyumunun oldukça iyi olduğunun bir göstergesidir.

Kestirilen model yardımıyla deneklerin sınıflandırma tablosu ise aşağıda verilmiştir. Sınıflandırma tablosu yapılırken kestirilen olasılık için kesim noktası 0.50 olarak belirlenmiştir ve böylece kestirilen olasılık değeri 0.50'nin altında olan denekler menopoza girmemiş, 0.50 ve üzerinde olan denekler ise menopoza girmiş olarak sınıflandırılmıştır.

Tablo 7. Final model kullanılarak yapılan sınıflandırma tablosu

Deneklerin menopoz evresinde olmasının		Beklenen frekansı		Doğru sınıflandırma yüzdeleri
		0	1	
Gözlenen frekansı	0	32	4	88,9
	1	4	75	94,9
Modeli doğru sınıflandırma yüzdesi				93,0

$$\text{Doğru sınıflandırma oranı} = \frac{(32 + 75)}{115} \times 100 = \%93.0 \text{ olarak bulunmuştur.}$$

### Sonuçlar ve Tartışma

Yapılan bu çalışma ile kadınlarda ortalama 45-55 yaşları arasında meydana gelen ve kadının doğurganlık döneminden kısırlık dönemine geçişi olarak tanımlanan menopoz evresine geçişi etkileyen faktörler bir model kavramı altında elde edilmeye çalışılmıştır. Bu amaç doğrultusunda menopoz şikayetleri ile Konya bölgesinde yer alan kırsal ve kent kesiminden özel bir kliniğe başvuran hastalar arasından rasgele 115 hasta kaydı incelenerek elde edilen verilere en uygun modelin belirlenmesi üzerine çalışılmıştır. Çalışmada temel olarak menopozun varlığı ve yokluğunun belirlenebilmesi için kurulacak lojistik model yardımıyla deneklerin bu fiziksel durumdan hangisine sahip olduklarını tahmin etmeye yardımcı olan bir sınıflandırma modeli elde edilmeye çalışılmıştır. Bu tahmini yapmadaki amaç hekimlerin hastalara koydukları tanılarının doğruluğunu istatistiksel olarak sınamaktır.

Öte yandan kurulan lojistik modelde değişken seçimine giderek, mümkün olduğunca az test ve tahlil bulgularını kullanıp hekime menopoz şikayetleri ile gelen hastanın hastalık grubunu en doğru şekilde tahmin edecek en iyi modelin bulunması amaçlanmıştır. Böylece hastalardan az sayıda bilgi alınmasının işlem kolaylığı sağlayıp zamandan kazanılacağı düşünülmüştür.

Yapılan bu çalışma ile menopoz evresine girmemiş olan 36 denekten 32'si kurulan model ile doğru olarak sınıflandırılmıştır. Menopozu girmiş olan 79 denekten ise 75'i model tarafından doğru olarak sınıflandırılmıştır. Yapılan bu çalışma neticesinde ise %3.48 oranında tahmini olarak yapılan menopoz teşhisi konmuş kadınların yanlış sınıflandırılması kadının halen doğurganlık özelliğine sahip olduğu ancak menstruasyon siklusunun yavaş yavaş düzensizleştiği menopoz öncesi döneme girmiş olmasından kaynaklanmaktadır. Bu safha premenopoz olup menopozu geçiş safhasında yer alan kadınlarında teşhis esnasında menopoz hastası olarak kayıtlara geçmesinden kaynaklanmaktadır.

115 hastaya ilişkin ele alınan verilere uygulanan lojistik regresyon analizi uyumlu sonuç vermiş ve %93 oranında verileri başarıyla sınıflandırmıştır.

### Teşekkürler

Yapılan bu çalışmada verilerinden yararlandığımız ve bizden desteğini esirgemeyerek yardımcı olan Dr. Sema Soysal Koç'a teşekkürlerimizi bir borç biliriz.

### Kaynaklar

- [1] İyit, N., **Lineer Olmayan Lojistik Regresyon Analizinde Model Kurma Stratejileri ve Bir Uygulaması**, Basılmamış Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya, (2003).
- [2] Bagley, S.C., White H., Golomb, B.A., **Logistic Regression in the Medical Literature: Standards for Use and Reporting, with Particular Attention to one Medical Domain**, Journal of Clinical Epidemiology, 54, 979-985, (2001).



- [3] Aldrich, H.J., Nelson, D.F., **Linear Probability Logit and Probit Models**, Sage Publications, London, (1986).
- [4] Hosmer, D. W. and Lemeshow, S., **Applied Logistic Regression**, John Wiley& Sons Inc. New York, (1989).
- [5] Menard S., **Applied Logistic Regression Analysis**, Sage Publications, California, (1995).
- [6] Kleinbaum, D.G., **Logistic Regression: A Self Learning Text**, Springer-Verlag, New York, (1994).
- [7] Berkson, J., **Application of the Logistic Function to Bio-Assay**, Journal of The American Statistical Association, 39, 357-365, (1944).
- [8] Cox, D.R., **Analysis of Binary Data**, Chapman and Hall, London, (1970).
- [9] Anderson, J.A., **Multivariate Logistic Compounds**, Biometrika, 66, 17-26, (1979).
- [10] Anderson, J.A., **Robust Inference Using Logistic Models**, Bulletin of International Statistical Institute, 48, 35-53, (1983).
- [11] Aranda-Ordaz, F.J., **On Two Families of Transformations to Additivity for Binary Response Data**, Biometrika 68, 357-363, (1981).
- [12] Brown, C.C., **On a Goodness of Fit Test for the Logistic Model Based on Score Statistics**, Communications in Statistics Theory and Method, 11, 1087-1105, (1982).
- [13] Johnson, W., **Influence Measures for Logistic Regression: Another Point of View**, Biometrika, 72, 59-65, (1985).
- [14] Pregibon, D., **Logistic Regression Diagnostics**, The Annals of Statistics, 9, 705-724, (1981).
- [15] Lesaffre, E., **Logistic Discriminant Analysis with Applications in Electrocardiography**, Doktora Tezi, Katholieke Universiteit Leuven, Belgium, (1986).
- [16] Lesaffre, E., Albert, A., **Multiple Group Logistic Regression Diagnostics**, Applied Statistics, 38, 425-440, (1989).
- [17] Bendel, R.B., and Afifi, A.A., **Comparison of Stopping Rules in Forward Regression**, Journal of the American Statistical Association, 72, 46-53, (1977).
- [18] Mickey, J., and Greenland, S., **A Study of the Impact of Confounder-Selection Criteria on Effect Estimation**, American Journal of Epidemiology, 129, 125-137, (1989).
- [19] Guerro, V.M. and Johnson, R.A., **Use of the Box-Cox Transformation with Binary Response Models**, Biometrika, 69, 309-314, (1982).
- [20] Hosmer, D. W., Lemeshow, S., **A Review of Goodness of Fit Statistics for Use in the Development of Logistic Regression Models**, Am. J. Epidemiol., 115, 92-106, (1982).

