

T. C. -
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

LOJİSTİK REGRESYON ANALİZİ
VE SAĞLIK BİLİMLERİNDE BİR UYGULAMA

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

YÜKSEK LİSANS TEZİ

107711

Cemil ÇOLAK
Biyostatistik Anabilim Dalı

T 107711

DANIŞMAN
Yrd. Doç. Dr. Saim YOLOĞLU

MALATYA - 2001

TEŐEKKÜR

Yüksek lisans eğitimin boyunca her türlü yardım ve desteğini esirgemeyen danışmanıcam Sayın Yrd. Doç. Dr. Saim YOLOĞLU'na, tezi hazırlama çalışmalarında bana her türlü yardımı sağlayan Sayın Araş. Gör. Atilla Halil ELHAN'a, Sayın Doç.Dr. Osman ARAÇBAŐI'na, Sayın Prof. Dr. Yıldır ATAKURT'a ve Sayın Araş. Gör. Dr. M. Cengiz OLAK'a teşekkürü bir borç bilirim.

Ayrıca bize eğitim imkanı veren devletimize, enstitü personelimize ve yetişmemem katkıları olan herkese çok teşekkür ederim.



İÇİNDEKİLER

1. GİRİŞ VE AMAÇ	1
2. GENEL BİLGİLER	3
2.1. Lojistik Regresyon nedir?	3
2.2. Neden Lojistik Regresyon popüler?	4
2.3. Lojistik Regresyon modeli	6
2.4. Modelin oluşturulması	8
2.4.1. Tahmin yöntemleri	8
2.4.1.1. En çok olabilirlik yöntemi	9
2.4.1.2. Yeniden ağırlıklandırılmış en küçük kareler yöntemi (RILS)	10
2.4.1.3. Minimum lojit Ki-kare yöntemi (MLCS)	10
2.4.1.4. Diğer tahmin yöntemleri	11
2.4.1.5. Tahmin yöntemlerinin karşılaştırılması	11
2.5. Tek değişkenli modelin önemlilik testi	12
2.6. Çok değişkenli lojistik regresyon analizi	15
2.6.1. Çoklu lojistik regresyonda en çok olabilirlik yöntemi	16
2.6.2. Çok değişkenli modelin önemlilik testi	18
2.7. Lojistik regresyonda sürekli ve kesikli bağımsız değişkenlerin kullanıldığı durum	19
2.8. Lojistik regresyon modelindeki katsayıların yorumlanması	20
2.8.1. Risk Ratio (oranının) ve Odds Ratio'nun (oranının) karşılaştırılması	24
2.8.2. Odds ratio (oranı) güven aralığının bulunması	25
2.9. Değişkenlerin etkileşimi (Interaction) ve etki karışımı (counfounding)	27
2.10. Değişken Seçimi	28
2.11. Çoklu Regresyonda Değişken Seçimi	33
2.12. Adımsal Teknikler ve Durdurma kuralları	34
2.12.1. İleriye doğru değişken seçim yöntemi	34
2.12.2. Geriye Doğru Değişken Çıkarma Yöntemi	35
2.13. Uyum İyiliği	36
2.13.1. D (Deviance) ve Pearson χ^2 (Ki-kare) Ölçütleri	37
2.13.2. Diğer uyum iyiliği ölçütleri	38

2.14. SPSS’de çoklu regresyon analizi yöntemleri	40
MATERYAL VE METOD	42
BULGULAR	46
TARTIŞMA	48
SONUÇLAR	70
ÖZET	71
SUMMARY	72
KAYNAKLAR	73



1. GİRİŞ VE AMAÇ

Çeşitli alanlarda yapılan bilimsel arařtırmalarda, özellikle sađlık bilimlerinde incelenen olayların karmařık ve olayı aıklamada kullanılabilen deđiřkenlerin de fazla sayıda olması nedeniyle, birden fazla deđiřkeni inceleyip, deđiřkenler arasındaki karmařık iliřkilerin yorumlanmasında kolaylıklar sađlayan ok deđiřkenli istatistiksel analizlerin kullanılması daha iyi sonu vermektir.

ok deđiřkenli istatistiksel analizlerle ilgili alıřmalar 1900'li yıllarda başlamıřtır. Bu analizde birbirleriyle iliřkili ok sayıda deđiřkenin oluřturduđu sistemin yapısı incelenmeye alıřılıp, daha basit bir forma indirgendikten sonra elde edilen bilgilerin yorumlanması, yorumlanması ve sonuta bir karara ulařmak iin kullanılmaktadır.(13)

Gözlemleri verilerin yapısında bulunan gruplara atamak iin kullanılan ok deđiřkenli istatistiksel yöntemler:

1. Kümeleme analizi
2. Diskriminant (Ayrırma) analizi
3. Lojistik regresyon analizidir

Kümeleme analizinde gözlemlerin atanacağı grup(küme) sayısı bilinmezken, lojistik regresyon ve diskriminant analizinde grup sayısı bilinmekte, mevcut bilgiler kullanılarak bir ayırimsama modeli elde edilmekte ve kurulan bu model yardımıyla veri yorumlanmasına eklenen yeni gözlemlerin gruplara atanması mümkün olabilmektedir.(10)

Sađlık bilimlerindeki arařtırmacıların inceledikleri konuyu etkileyen birden ok etken bulunmaktadır. Bu etkenlerin ayrı ayrı bađımlı(sonu) deđiřken üzerine olan etkilerine bakılarak hepsinin birlikte yapmış olduđu etkiyi de arařtırmak istemektedirler. Bađımlı deđiřkenin yapısı kesikli, bađımsız deđiřkenlerin yapıları genellikle sürekli ve kesikli karıřımı içermektedir. Ayrıca belirli özelliklere sahip bir bireyin herhangi bir hastalıđa yakalanma olasılığı hesaplanabilmektedir. Böyle durumlarda lojistik regresyon analizi kullanılmaktadır.(5)

Son yıllarda yapılan tıbbi arařtırmalarda lojistik regresyonun analizinin sık olarak kullanıldığını görülmektedir. Epidemiyolojik arařtırmalarda, incelenen hastalıđa neden olan etkenler veya faktörler (aıklayıcı deđiřkenler) ile aıklanan (bađımlı) deđiřken arasındaki iliřkinin ne olduđu önemli bir konudur.

Bir örnek verecek olursak: iki gruplu hastalık değişkeni olarak koroner arter taliğini ele alalım. Bir bireyin koroner arter hastası olduğunu 1 değeri, 0 ise koroner arter taşı olmadığını temsil etsin. Hastalıkla ilişkili değişkenler ise; düzenli egzersiz yapma, sigara içip içmeme, cinsiyet, yaş, ... olduğunu kabul edelim. Açıklayıcı değişkenler açıklanan değişken arasındaki ilişkiyi açıklamak ve hastalığı önceden tahmin etmek için kullanılan yöntemlerden birisi lojistik regresyon analizidir.

Lojistik regresyon analizi konusundaki çalışma ve uygulamalar daha çok son ara dayanmaktadır. Yöntem çeşitli varsayım bozulmaları durumunda (normallik, ortak varyansa sahip olma gibi) diskriminant analizi ve çapraz tablolara bir alternatif olarak lanılmaktadır. Ayrıca bağımlı değişkenin 0,1 gibi ikili (Binary) ya da ikiden çok düzey ren kesikli değişken olması durumunda normallik varsayımının bozulması nedeniyle de ğrusal regresyon analizine alternatif olmaktadır. Alternatif bir yöntem olarak gelişen ve ıelleştirilmiş lineer modellerin (GLM) özel bir durumu olan lojistik regresyon analizi tıpta, demiyolojide, meteorolojide, ekonomide,...vb. sıkça kullanılmaktadır.(6)

Çözümlemede elde edilen modelin matematiksel olarak esnek ve kolay umlanabilir olması, ayrıca anlamlı yorumlara götürebilmesi, yönteme olan eğilimin nasında önemli faktörlerdir.

Ayrıca lojistik regresyon analizinde, Epidemiyoloji bilimindeki vaka kontrol ştırmalarında odds ratios ve cohort araştırmalarında relative risk daha az matematiksel mlerle elde edilip, yorumlanabilmektedir. Sağlık bilimlerinde karşılaşılan hastalıklara en bilgi yardımıyla kısa sürede ve doğru olarak tanı koyma imkanını verebilen yöntemlerden isi lojistik regresyon analizidir.(4)

Bu çalışmada hem kesikli hem de sürekli değişkenleri içeren kardiyolojik verilere stik regresyon analizi uygulanarak, ileriye doğru değişken seçim yöntemi ile geriye doğru işken eleme yönteminin sonuçları karşılaştırılacak ve en iyi lojistik model bulunmaya ışıacaktır. Ayrıca en sık karşılaşılan risk faktörleri tespit edilip, en az bilgi ile kısa sürede talığı teşhis etme yolları aranacaktır.

2. GENEL BİLGİLER

2.1. Lojistik Regresyon nedir?

Epidemiyolojik arařtırmalarda çok sık olarak karřılařılan çok deęiřkenli problemi klamakla bařlayalım. Arařtırmacıların merak ettikleri önemli sorulardan birisi: cevap aęımlı) deęiřkeni (hastalık gibi) ile aıklayıcı (baęımsız) deęiřkenler arasındaki iliřkinin ne luęudur. Mesela: iki gruplu hastalık deęiřkeni olarak Koroner Arter hastalıęını ele alalım. 1 bireyin hasta, 0 ile hasta olmadığı gsterilsin. Evet ve Hayır řeklinde kodlanan sigara iip neme durumu ile Koroner Arter hastalıęı (KAH) arasındaki iliřkinin boyutunu incelemek enilsin. Ayrıca sigara etkenine ek olarak yař, cinsiyet,... gibi faktörler ise kontrol ğiřkenleri olarak adlandırılmaktadır. Sigara, yař, cinsiyet,... ile bunların kombinasyonları; ğara ile yař, sigara ile cinsiyet,... gibi deęiřkenlerin oluřturduęu topluluęun yardımıyla vap(hastalık) deęiřkenini aıklamak veya tahmin etmek isteyelim. Byle bir duruma 'çok ğiřkenli problem adı verilip, bu amaca yönelik olarak kullanılan çok deęiřkenli istatistiksel ntemlerden biri olan Lojistik Regresyon analizi; birden çok baęımsız deęiřkenlerin, bir otomous(ikili) baęımlı deęiřken ile iliřkisini aıklamak iin kullanılan bir matematiksel model yaklařımıdır.(2)

Lojistik regresyon analizi; cevap deęiřkeninin kategorik, ikili (binary, dichotomous), lü ve çoklu kategorilerde gözlendięi durumlarda aıklayıcı deęiřkenlerle neden-sonuç řkisini belirlemede yararlanılan bir yöntemdir. Aıklayıcı deęiřkenlere göre cevap ğiřkeninin beklenen deęerleri olasılık olarak elde edildięi bir regresyon yöntemidir. Ayrıca aflama ve atama iřlemi yapmaya yardımcı olan bir regresyon yöntemidir.(3)

Lojistik regresyon analizi, temelde regresyon analizi olmakla birlikte bir ayırıcı analiz ma özellięini de tařımaktadır. Ancak lojistik regresyon analizi, baęımsız deęiřkenin yapısı e kombinasyonu yönünden diskriminant analizinden farklılık göstermektedir.

Lojistik regresyon analizinin regresyon analizinden ise üç önemli farklılığı vardır. Bunlar:

1. Regresyon analizinde bağımlı değişken sayısal iken, lojistik regresyon analizinde kesikli bir değer olmalıdır.
2. Regresyon analizinde bağımlı değişkenin değeri, lojistik regresyonda ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir.
3. Regresyon analizinde bağımsız değişkenlerin çoklu normal dağılım göstermesi şartı aranırken, lojistik regresyon analizinin uygulanabilmesi için bağımsız değişkenlerin dağılımına ilişkin hiç bir kısıtlama getirmez.(12)

2. Neden Lojistik Regresyon popüler?

Lojistik regresyonun güncel olmasının nedenleri şunlardır:

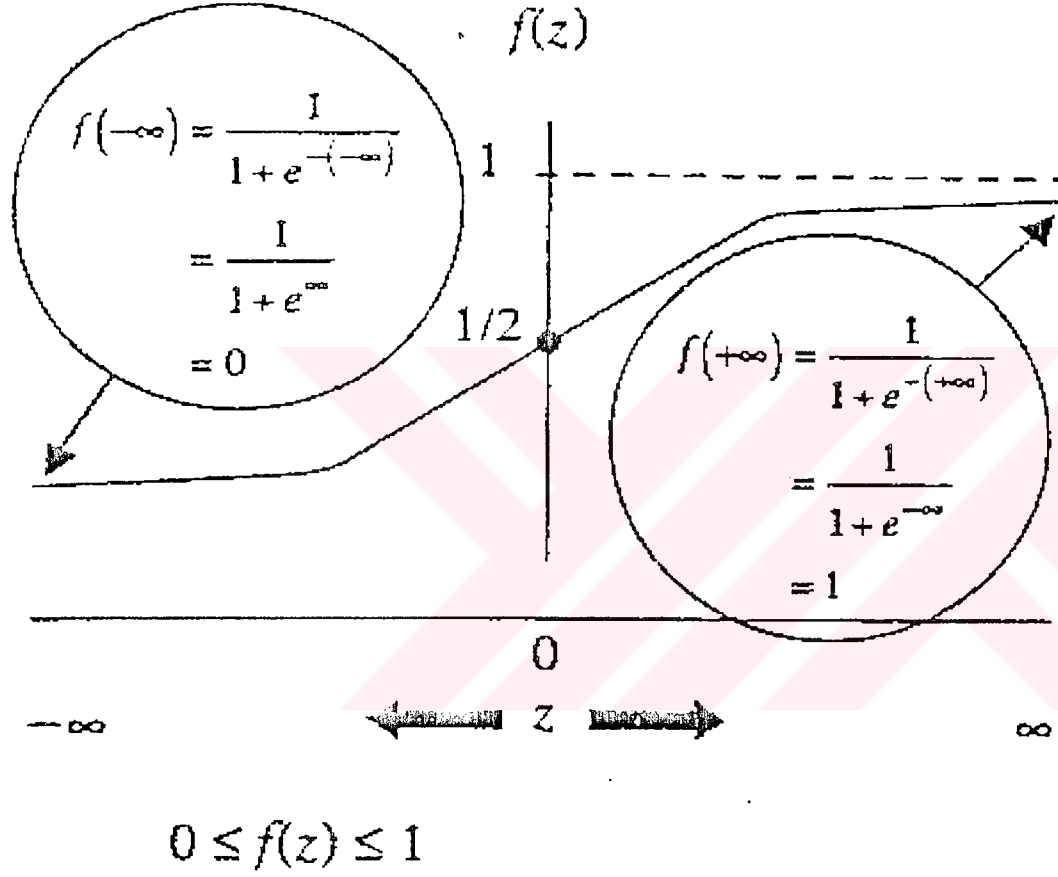
1. Cevap değişkeni kesikli iken, açıklayıcı değişkenlerin hem kesikli hem de sürekli olduğu durumlarda uygulanabilmektedir.
2. Lojistik modelin parametreleri Epidemiyoloji'de yapılan ölçümlere benzediği için yorumları kolay olmaktadır. Epidemiyoloji'de katsayıların exponansiyeli(üssü) hastalık riski olarak yorumlanabilmektedir.
3. Lojistik modelin parametre sayısına karşılık gelen doğrusal regresyon modeli ve diskriminant fonksiyonu ile aynı olmaktadır.
4. Lojistik modele dayalı analizler için çok sayıda paket programlar vardır.
5. Açıklayıcı değişkenlerin olasılık fonksiyonlarının dağılımı üzerinde kısıt olmaması(yarı parametrik) nedeniyle çeşitli testler uygulanabilmektedir. Epidemiyoloji'de, tıpta, deneysel verilerin analizinde, meteoroloji'de,...vb. alanlarda sıkça kullanılan lojistik regresyon analizi farklı varsayımlar durumunda aynı lojistik formülasyona götürdüğü için varsayım bozulmalarına daha güçlü bir yöntem olmaktadır.(4)
6. Matematiksel olarak kolay olup, anlamlı sonuçlar vermektedir.
7. Öncelikle lojistik regresyon analizinin dayalı olduğu matematiksel formu açıklayan lojistik fonksiyonu inceleyelim. Bu fonksiyon $f(z)$ ile gösterilsin.

$$f(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} \quad (1)$$

$-\infty < z < \infty$ değer almaktadır. $z = -\infty$ iken $f(z)=0$ ve $z = \infty$ iken $f(z)=1$ olmaktadır. Yani her z için $0 < f(z) < 1$ 'dir. Bu model, 0 ile 1 arasındaki bir ihtimali açıklamak için kullanılmaktadır. Epidemiyoloji'de böyle bir ihtimal bir bireyin hastalığa yakalanma riskini vermektedir.

Lojistik modelin S-şeklinde olması epidemiyologların ilgisini çekmektedir. Buradaki birden fazla risk faktöründen oluşan bir dizin olması nedeniyle, S-şeklindeki model, çok değişkenli epidemiyolojik araştırmalarda sıkça kullanılmaktadır.(2)

Şekil 1. Lojistik Fonksiyonu



8. Diskriminant analizinin iki temel varsayımı olan; verilerin çok deęişkenli normal ağılıma sahip olması ve grupların ortak varyans-kovaryans matrisli olmaları genellikle ağılanamadığı için bu yöntemin kullanımı kısıtlı olmaktadır. Varsayımların bozulması durumunda kullanılan bazı özel diskriminant fonksiyonları olmakla birlikte yorum kolaylığı ve kestirim güçlülüğü gibi nedenlerden dolayı lojistik regresyon analizi son yıllarda daha çok kullanılmaktadır.(10)

2.3. Lojistik Regresyon modeli

Regresyon problemlerinde anahtar deęer, verilen bir bağımsız deęişkenin deęerine bağılı olarak bağımlı (sonuç) deęişkenin ortalama deęerini bulmaktır. Bu deęer koşullu ortalama olarak adlandırılır ve $E(Y|x)$ ile gösterilir. $E(Y|x)$; verilen bir x deęeri için Y 'nin beklenen deęeri olarak ifade edilir. Burada Y ; bağımlı deęişkeni, x ise, bağımsız deęişkeni ifade etmektedir. Lineer regresyon analizinde, koşullu ortalamanın x 'in lineer bir denklemi olduđu varsayılır;

$$E(Y|x) = \beta_0 + \beta_1 x \quad ; \quad -\infty < x < +\infty \quad (2)$$

burada, x 'in aralığının $-\infty$ ile $+\infty$ arasında deęişmesi nedeniyle, $E(Y|x)$ 'in mümkün olan her deęeri alabilmektedir. Lojistik regresyon analizinde ise, koşullu ortalama 0 ile 1 arasında deęer almak zorundadır.

$$0 \leq E(Y|x) \leq 1$$

Lojistik regresyon analizinde, $E(Y|x) = \beta_0 + \beta_1 x$ eşitliğinin sol tarafı 0-1 arasında sınırlı olasılık deęerleri aldığından ve bu deęerler sonsuz deęerler alabilen bağımsız (açıklayıcı) deęişkenlerle ilişkilendirildiğinden, söz konusu eşitlik her zaman sağlanamamaktadır. Böyle bir durumda en iyi çözüm; sonuç deęeri olarak ifade edilen olasılık deęerinin çeşitli dönüşümlerle $-\infty$ ile $+\infty$ arasında tanımlı hale getirilmelidir.

İki düzey içeren bir sonuç deęişkeninin analizinde kullanılmak için önerilen bir çok dağılım fonksiyonu vardır. En yaygın kullanılanlardan iki tanesi: lojit ve probit dönüşümleridir.

Lojistik dağılımı seçmek için iki tane önemli neden vardır. İlk neden: lojistik regresyon analizinde varsayım kısıtlaması olmaması nedeniyle kullanım rahatlığının yanı sıra, naliz sonucu elde edilen modelin matematiksel olarak çok esnek olması, ikinci neden ise; iyolojik olarak kolay yorumlanabilir olmasıdır. Lojistik dağılım kullanıldığı zaman, x ilindiğinde Y 'nin koşullu ortalamasını göstermek için $\pi(x) = E(Y|x)$ ile ifade edelim. Lojistik regresyon modeli aşağıdaki gibidir:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3)$$

Lojistik regresyon çalışmasına merkez olacak $\pi(x)$ 'in bir dönüşümü yukarıda ahsedildiği gibi lojit dönüşümüdür. Bu dönüşüm $\pi(x)$ cinsinden aşağıdaki gibi tanımlanır:

$$g(x) = \ln[\pi(x)/(1-\pi(x))] = \beta_0 + \beta_1 x \quad (4)$$

$\pi(x)$ başarı, $1-\pi(x)$ başarısızlık ihtimali olarak düşünülürse, lojistik model; başarı ihtimalinin başarısızlık ihtimaline oranının logaritması olarak tanımlanabilir. Lojistik modelin ağımlı değişkenini oluşturan lojit dönüşümün ($\text{Lojit}(\pi(x)) = \ln[(\pi(x))/(1-\pi(x))]$) aşica özellikleri şunlardır:

1. $\pi(x)$ arttıkça, lojit ($\pi(x)$)'de artar.
2. $\pi(x)$, 0-1 arasında değer alırken, lojit ($\pi(x)$) bütün gerçel doğru üzerinde değer alabilmektedir.
3. Eğer $\pi(x) < 0,5$ ise, lojit ($\pi(x)$) < 0 ve $\pi(x) > 0,5$ ise lojit ($\pi(x)$) > 0 olur. Gözlemler sınıflara en basit şekilde bu kurala göre atanabilir. (6)

Bu dönüşümün önemi $g(x)$ 'in lineer regresyon modelinin istenen bütün özelliklerini şımasıdır. Lojit $g(x)$ parametreleri bakımından lineer (doğrusal), sürekli ve x 'in aldığı gerlere bağlı olarak $-\infty$ ile $+\infty$ arasında değişebilmektedir.

neer ve lojistik regresyon arasındaki farklardan birisi de, sonuç değişkeninin dağılımıyla ilidir. Lineer regresyonda sonuç değişkeninin bir gözlemi $y = E(Y|x) + \varepsilon$ 'dir.

hata terimi olup, gözlemin koşullu olasılıktan sapma miktarını göstererek, 0 ortalama ve σ^2 varyanslı normal dağılıma sahiptir. İki düzeyli sonuç değişkenlerinde ise, x rildiğinde sonuç değişkeninin değeri $y = \pi(x) + \varepsilon$ 'dir.

Eğer $y=1$ ise $\pi(x)$ ihtimalle $\varepsilon=1-\pi(x)$, $y=0$ ise $1-\pi(x)$ ihtimalle $\varepsilon=\pi(x)$ olup, ε sıfır ortalama ve $\pi(x)(1-\pi(x))$ varyanslı binom dağılımına sahiptir.

Özet olarak; sonuç değişkeni iki düzeyli olduğu zaman regresyon analizinde;

1. Regresyon denkleminin koşullu ortalamasının 0 ve 1 arasında sınırlandırılarak yazılması gerekmektedir. Lojistik regresyon analizini bu kısıtlamayı sağlamaktadır.
2. Hataların dağılımı, normal dağılım yerine binom dağılımı olması ve analizin bu temel üzerine dayanmasıdır.
3. Lineer regresyon analizinde kullanılan prensipler, lojistik regresyon analizinde de kullanılmaktadır.(1),(12)

2.4. Modelin oluşturulması

Hastalık değişkeni olarak, 0 ve 1 olarak kodlanan KAH'nı alalım. 0; hastalığın yokluğunu, 1 ise varlığını gösterebilir. Bir grup birey üzerinde çeşitli ölçümler yapılsın. Örneğin: X_1 ; bireylerin sigara içip içmedikleri, X_2 ; Yaş ve X_3 ; cinsiyet olsun. Bu bireyler Cohort çalışmalarında olduğu gibi belirli bir zaman gözlemlensin. X_1, X_2, X_3 gibi bağımsız değişkenlere sahip hasta olmayan bireylerin belirli bir süre boyunca izlenmesiyle, hastalığın ortaya çıkma ihtimali açıklanmak istensin. Koşullu ihtimal $P(D=1 \mid X_1, X_2, X_3)$ ile gösterilsin. Burada lojistik model:

$$P(D=1 \mid X_1, X_2, X_3) = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)}}$$

Burada D ; hastalık değişkenini yani bağımlı değişkeni ifade etmektedir. α ve β_i 'ler bilinmeyen parametreler olup, bu veriler yardımıyla tahmin edilecektir. Şimdi bilinmeyen parametreleri tahmin etmek için kullanılan tahmin yöntemlerini inceleyeceğiz.

2.4.1. Tahmin yöntemleri

İki gruplu bir lojistik modelin katsayılarının tahmin edilmesinde kullanılan yöntemler:

1. En Çok Olabilirlik yöntemi (Maximum Likelihood)
2. Yeniden ağırlıklandırılmış en küçük kareler yöntemi (RILS)
3. Minimum lojit ki-kare yöntemi (MLCS)

2.4.1.1. En çok olabilirlik yöntemi

En çok olabilirlik yöntemi bilinmeyen parametrelerin tahmin edilmesinde kullanılan yöntemlerden biridir. Lineer regresyonda ise parametre tahminlerinde en küçük kareler yöntemi kullanılmaktadır. En çok olabilirlik yöntemi ve en küçük kareler yöntemi, bağımlı değişkenin normal dağılıma sahip olması şartıyla, klasik lineer regresyonda aynı sonucu veren farklı metodlardır. En çok olabilirlik yönteminde, kompleks hesaplamaların yapılması gerektiğinden bu işlemleri yapabilecek paket programları önceleri yok idi. Bu nedenle sıkça kullanılmamaktaydı. Ancak son yıllarda bir çok paket programların yazılması ile sıklıkla kullanılmaktadır. En çok olabilirlik yöntemi, hem lineer olmayan hem de lineer modellerin tahmin edilmesinde kullanılabilir. Lojistik model, lineer olmayan bir model olduğundan lojistik regresyonda tahmin yöntemi olarak en çok olabilirlik yöntemi tercih edilmektedir. En çok olabilirlik yöntemi, gözlenen veri kümesini elde etmenin olasılığını maksimum yapan bilinmeyen parametrelerin değerlerini verir. Bu metodu uygulamak için önce en çok olabilirlik fonksiyonunun oluşturulması gerekmektedir. Bu parametrelerin en çok olabilirlik tahminleri, fonksiyonu maksimum yapacak şekilde seçilir. Böylece tahminler gözlenen değerlere en yakın olan değerler olarak seçilir. (8)

Bilinmeyen parametrelerin tahmin edilmesinde karmaşık matematiksel işlemlerin yapılması ve SPSS gibi paket programların bu işlemleri kolayca yapması nedeniyle, matematiksel işlemler üzerinde durulmayacaktır.

Lojistik regresyonda parametrelerin tahmin edilmesinde kullanılan alternatif iki en çok olabilirlik yaklaşımı vardır. Bunlar: Koşulsuz (Unconditional) ve koşullu (Conditional) en çok olabilirlik yöntemleridir. Lojistik regresyonu kullanacak bir araştırmacı önce bu iki yöntemden hangisinin verilerine uygun olacağına karar vermelidir. Koşulsuz metod için kullanılabilecek paket programlar: SPSS, EGRET, SAS(LOGIST),... gibidir. Koşullu metod için kullanılabilecek paket programlar: SAS(PECAN), SPIDA,... gibi programlardır. Peki araştırmacı bu yöntemler arasında nasıl bir seçim yapacaktır?

Eğer modeldeki parametre sayısı denek sayısına göre az ise, bu durumda koşulsuz metod, modeldeki parametre sayısı denek sayısına göre fazla ise, bu durumda koşullu metod uygulanır. Mesela; parametre sayımızın 13 ve denek sayımızın 600 olduğunu kabul edelim. Bu durumda parametre sayısı denek sayısına göre oldukça küçük olduğundan koşulsuz EÇÖ metodu kullanılır. Eğer parametre sayımız 110 ve denek sayımız 200 ise parametre sayısı denek sayısına göre fazla olduğundan koşullu EÇÖ metodu kullanılır.

Sonuç olarak, arařtırmacı zaman ve paradan tasarruf yapmak istiyorsa ve modeli řulsuz EÇO metoduna uygun ise, kořulsuz metodu kullanmalıdır. Buna karřın kořullu ÇO metodu daima yansız sonuçlar verirken, kořulsuz EÇO metodu uygun olmayan bazı rumlarda yanlı sonuçlar verebilmektedir. Bir kararsızlık söz konusu olduęunda, kořullu ÇO metodu kullanılması önerilmektedir. (2)

2.4.1.2. Yeniden aęırlıklandırılmıř en küçük kareler yöntemi (RILS)

Gruplandırılmıř verilerde J grubun her birinde n_j denemeden r_j başarı elde edilsin. $1,2,\dots,J$ için başarı oranı $P_j = r_j/n_j$ olarak tanımlanmaktadır. $\text{Var}(r_j/n_j) = P_j(1-P_j)/n_j$ olduęundan r binom daęılımlı gözlem için varyans deęiřmektedir. Bu durumda Logit (r_j/n_j) 'nin ıklayıcı deęiřkenler üzerinde $W_j = n_j / P_j(1-P_j)$ aęırlığı ile aęırlıklandırılmıř regresyonu daha uygun olacaktır. W_j aęırlık deęerleri de P_j 'nin bir fonksiyonu olduęu için en küçük kareler yöntemi iteratif olarak uygulanacak, her adımda aęırlıklar tahminlere baęlı olarak yeniden de edilecektir.(10)

2.4.1.3. Minimum lojit Ki-kare yöntemi (MLCS)

Log-doęrusal modeli test etmede kullanılan ve aęırlıklı en küçük kareler tahmininin özel bir türü olarak $2 \times J$ çapraz tablolarında Berkson (1955) tarafından önerilen yöntemde, ıklenen ve gözlenen lojit deęerleri arasındaki farktan yararlanılmaktadır.

Yöntemin lojistik regresyonda kullanımı tekrarlı veriler olduęu durumlarda geçerlidir. Yeniden aęırlıklandırılmıř iteratif en küçük kareler yönteminde sözü edilen P_i deęeri:

$$P_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad \text{gibidir.} \quad (5)$$

Bu olasılık üzerinde yapılan lojit dönüşüm minimum lojit ki-kare yönteminde baęımlı deęiřkeni oluřturmaktadır. Tahminde kullanılacak aęırlık deęeri $n_j P_j(1-P_j)$ olarak elde edilmektedir. Yöntem, lojit deęeri olarak tanımlanan baęımlı deęiřkenin baęımsız deęiřkenler üzerinde aęırlık deęeri ile aęırlıklandırılmıř regresyondan en küçük kareler tahminlerini elde etmeye dayanır. Tek adımda bulunan aęırlıklı en küçük kareler tahminleri minimum lojit ki-kare adını almaktadır. Olasılık deęerinin 0 ya da 1 olduęu durumda lojit deęeri tanımlı olmayacağı için P_j yerine $P_j + 1/2n_j$ deęerinin konulduęu ayarlanmıř lojit ki-kare yöntemi kullanılmaktadır.(6)

2.4.1.4. Diğer tahmin yöntemleri

ukarıdaki tahmin yöntemlerine ek olarak kullanılan diğer tahmin yöntemleri:

1. İteratif olmayan en küçük kareler yöntemi
2. Diskriminant fonksiyonu yöntemidir.

İteratif olmayan en küçük kareler yönteminin getirdiği bir kısıt, veri kümesindeki çoğu gözlem için olasılık tahmin değerinin sıfırdan farklı olmasıdır. Bu kısıt açıklayıcı değişken sayısı çok olduğunda sağlanamayacağı için yöntem, tek açıklayıcı değişken durumu dışında kullanışlı değildir.

Diskriminant fonksiyonuna bağlı tahmin açıklayıcı değişkenlerin yanıt değişkeni koşulu altında normal dağıldığı varsayımına dayanmaktadır. Varsayım bozulduğunda, elde edilen diskriminant fonksiyonun tahminleri yanlı olacağından, yöntem bir ön analiz olarak ve lojistik modele başlangıç tahminlerini vermek amacıyla kullanılabilir.(4)

2.4.1.5. Tahmin yöntemlerinin karşılaştırılması

Lojistik modele dayalı analizlerde parametre tahmininde kullanılan yöntemlerden En Çok Olabilirlik yöntemi her zaman tutarlı, etkin ve yeterli tahminler verir. Tahminler ise her zaman yansız olmayıp, asimtotik olarak yansızdır. Ayrıca bu tahminler normal dağılıma sahiptir. Doğrusal olasılık modelinin ağırlıklı en küçük kareler tahminleri ile lojistik modelin çok olabilirlik tahminleri benzer istatistiksel özelliklere sahiptir. Varsayımlar sağlandığında en küçük kareler ve en çok olabilirlik tahminleri aynı özellikleri gösterir; ancak bir fark vardır. Bu fark en çok olabilirlik yönteminin olabilirlik denkleminin doğrusal olmayıp iteratif türevler ile sonuca gitmesidir. Bu durum hesaplama maliyetini arttırıp, çok zaman almaktadır.

Öte yandan minimum lojit ki-kare yönteminden de asimtotik olarak etkin ve yeterli tahminler elde edilmektedir. Bu iki yöntemin ortak özelliği; yansız, etkin ve normal tahminler sağlamalarıdır. Sonuç olarak, nokta tahmini için minimum lojit ki-kare yöntemini, çıkarsama için ise en çok olabilirlik yönteminin kullanılması önerilmektedir.(10)

2.5. Tek deęişkenli modelin önemlilik testi

Katsayılar tahmin edildikten sonra, modeldeki deęişkenlerin önemli olup olmadıklarının incelenmesi gereklidir. Bu test modeldeki bağımsız deęişkenlerin bağımlı deęişkenle olan ilişkilerinin önemli olup olmadığının testi şeklindedir.

Modelde tek bağımsız deęişken olduğunda durumda deęişkenin önemlilik kontrolü şu şekildedir.

İlk olarak deęişkenin önemliliğinin araştırılması konusunda aklımıza gelecek ilk soru; önemliliği araştırılan bağımsız deęişken, bağımlı deęişkeni açıklamada ne kadar etkin olduğudur” veya şu şekilde de ifade edilebilir; “ inceleme altındaki deęişkeni içeren model, sonuç deęişkeni açısından incelenen deęişkeni içermeyen modelden daha fazla bilgi verebiliyor mu?” Bu sorunun cevabı, bağımlı deęişkenin gözlenen deęerlerini, her iki modelden tahmin edilen deęerlerle karşılaştırılarak bulunabilir. Eğer incelenen bağımsız deęişkeni içeren model, o deęişkeni içermeyen modelden daha iyi ise, (bağımlı deęişken hakkında daha fazla bilgi veriyorsa) o deęişkenin önemli olduğu söylenebilir.

Lojistik regresyonda katsayıların önem testi için ana prensip sorgulama altındaki deęişkeni içeren ve içermeyen modellerden elde edilen tahmin deęerlerinin sonuç deęişkeninin gözlenen deęerleriyle karşılaştırılmasıdır. Gözlenen ve tahmin edilen karşılaştırma işlemi kesiklilik nedeniyle log-olabilirlik (log-benzerlik) fonksiyonu ile yapılır. Olabilirlik fonksiyonlarını kullanarak gözlenen ve tahmin edilen deęerleri karşılaştırmak aşağıdaki ifade ile yapılmaktadır.

$$D = -2 \ln \left[\frac{\text{Şu andaki modelin olabilirlięi}}{\text{Doymuş modelin olabilirlięi}} \right] \quad (6)$$

Büyük parantezin içindeki ifadeye olabilirlik oranı denilmektedir. D istatistiğine Deviance adı da verilmektedir. Logaritmanın eksi iki (-2) katının alınmış olmasının matematiksel olarak bir anlamı olduğu kadar dağılımı bilinen bir deęer de elde edilebilmektedir. Bu elde edilen deęer hipotez testi olarak kullanılmaktadır. Böyle bir teste, olabilirlik oran testi adı verilmektedir. Uyum iyiliğinde D ölçütü oldukça önemlidir. Bağımsız bir deęişkenin önemine karar vermek için denklemden bağımsız deęişkenin olduğu ve olmadığı durumlardaki D deęerleri karşılaştırılır.

Bağımsız değişkeni içermesinden dolayı ortaya çıkan D'deki değişim aşağıdaki gibidir:

$$G = D (\text{Değişkensiz Model}) - D (\text{Değişkenli Model}) \quad (7)$$

Bu istatistik doğrusal regresyonda kullanılan F testindeki pay kısmı ile aynı rolü üstlenir.

$$G = -2 \ln \left[\frac{\text{Değişkensiz modelin olabilirliği}}{\text{Değişkenli modelin olabilirliği}} \right] \quad (8)$$

Bütün değişkenleri içeren model ile tahmin edilen modele ilişkin olabilirlik oranlarının farkına dayanan ölçütlerin ki-kare dağılımı göstereceği düşüncesinden hareketle rulan modelin geçerliliği sınanmaktadır. Bu yolla modele girecek bağımsız değişkenlere karar verilmektedir. $\beta_1 = 0$ hipotezi altında, G istatistiği 1 serbestlik dereceli ki-kare dağılımı sterecektir.(5)

Log olabilirlik ve olabilirlik oran testinin hesaplamalarını kolayca yapan bir çok paket programı mevcuttur. Bu paket programlarından SPSS kolaylıkla bu testlerin sonuçlarını görmektedir. Bu şekilde modele giren her değişkenin önem kontrolünün hızlıca yapılmasına imkan sağlanır. Tek bir bağımsız değişkenin olduğu durumda, ilk olarak sabit terimi içeren model kurulur. Sonra sabit terimle birlikte bağımsız değişkenin içinde bulunduğu model oluşturulur. Elde edilen iki log olabilirlikten olabilirlik oran testi hesaplanıp değişkenin önemli olup olmadığı karar verilir.

Yine değişkenlerin önemlilik testinde kullanılan diğer bir test ise, Wald testidir. Wald istatistiği, aşağıdaki şekilde tanımlanır;

$$W = \hat{\beta}_1 / SE(\hat{\beta}_1) \quad (9)$$

Burada $H_0 : \beta_1 = 0$ hipotezi test edilmekte olup, bulunan W deęeri H_0 hipotezi altında standart normal daęılım göstermektedir. Hauck ve Donner (1977) Wald testinin etkinlięini incelemişler ve sonuçta olabilirlik oran testinin kullanılmasını önermektedir.

Dięer bir test ise Score testidir. Bu testin hesaplamasında matematiksel işlemlerin az olması bir avantajı iken, bir çok paket programı tarafından hesaplanamaması bir dezavantajdır. Score testi de Wald testi gibi standart normal daęılım göstermektedir. Score testi aşağıdaki formül ile hesaplanır.

$$ST = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sqrt{\bar{y} (1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (10)$$

Sonuç olarak, ikili sonuç deęişkeni olduęu durumlarda, olabilirlik oran testinin kullanılması tavsiye edilmektedir.(1)

2.6. Çok deęişkenli lojistik regresyon analizi

x_1, x_2, \dots, x_p gibi p tane bağımsız deęişkenin olduğunu ve bu deęişkenlerin her birisinin en az aralık ölçekli olduğunu kabul edelim. Bağımlı deęişkenin var olduğu zaman ($Y=1$) koşullu olasılık $P(Y=1 \mid x) = \pi(x)$ 'dir. Çok deęişkenli lojistik modelin lojiti aşağıdaki gibidir:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad \text{olarak bulunur.} \quad (11)$$

Eđer bazı bağımsız deęişkenler kesikli, nominal ölçekli (ırk, cinsiyet, v.b.) ise, o zaman bu deęişkenleri sürekli deęişkenlermiş gibi denkleme dahil etmek yanlış olacaktır. Çünkü bu deęişkenlere verilen sayıların herhangi bir sayısal deęerleri yoktur. Bu durumda çeşitli dizayn deęişkenlerinin ya da dummy deęerlerinin (kukla deęişkenlerinin), kategorik olan bu deęişkenleri temsil etmesi için kullanılması gereklidir.

Bir örnek verecek olursak; ırk deęişkeni: beyaz, siyah ve dięerleri olarak kodlansın. Kategori sayısı 3 olduğundan 2 tane dizayn deęişken (D_1 ve D_2) kullanılmalıdır. Kodlama metotlarından ve SPSS'de de kullanılan birisi, referans kategori olarak alındığında aşağıdaki gibi kodlama yapılmıştır.

Tablo 1. Üç kategorili bir deęişkenin dizayn deęişkenlerinin kullanılmasına bir örnek

IRK	D_1	D_2
Beyaz	0	0
Siyah	1	0
Dięer	0	1

Eđer nominal bir deęişkenin k kategorisi varsa, o zaman $k-1$ tane dizayn (kukla) deęişkeni kullanılmalıdır. j .inci bağımsız deęişken (x_j)'nin k_j tane kategorisi varsa, k_j-1 dizayn deęişkeni D_{ju} olarak ve $u=1, 2, \dots, k_j-1$ için β_{ju} ile katsayılar gösterilsin. j .nci deęişkeni kesikli olan p deęişkeni olan model için lojit aşağıdaki gibidir:

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \sum_{u=1}^{k_j-1} \beta_{ju} D_{ju} + \beta_p x_p \quad (12)$$

ile ifade edilir.(8)

2.6.1. Çoklu lojistik regresyonda en çok olabilirlik yöntemi

Birbirinden bağımsız n tane (x_i, y_i) , $i=1,2,\dots,n$ değişkeninin olduğunu kabul edelim. Tek değişkenli modelde olduğu gibi modelin kurulması için $\beta_0, \beta_1, \dots, \beta_p$ gibi katsayıların oluşturduğu tahmin vektörünü elde etmemiz gerekir. Tek değişkenli durumda olduğu gibi tahmin yöntemimiz maksimum olabilirlik olacaktır. Burada log olabilirlik fonksiyonunun $p+1$ katsayıya göre türevi alınıp, sıfıra eşitlenerek $p+1$ tane olabilirlik denklemi elde edilir.

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (13)$$

ve

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0, \quad j=1,2,\dots,p \quad (14)$$

Bu denklemler çözüldüğünde $\hat{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ parametreleri bulunur. Tek değişkenli modeldeki gibi bu denklemlerin çözümünü yapabilen bir çok paket programı mevcuttur. Burada tahmin edilen değerler $\hat{\pi}(x_i)$ 'dir.

$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$ denklemindeki $\pi(x)$ ifadesinin değeri, $\hat{\beta}$ ve x_i kullanılarak bulunmuştur.

Tahmin edilen katsayıların varyans-kovaryans değerleri, log olabilirlik fonksiyonlarının ikinci dereceden kısmi türevlerinden oluşan bir matristen elde edilir. Bu türevlerin genel şekli aşağıdaki gibidir:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (15)$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_u} = - \sum_{i=1}^n x_{ij} x_{iu} \pi_i (1 - \pi_i) \quad (16)$$

Burada $j, u = 0, 1, \dots, p$ 'dir. (15) ve (16) eşitliklerinde verilen ifadelerin negatiflerini içeren $(p+1) \times (p+1)$ boyutlu bir matris $I(\beta)$ ile gösterilip, bilgi (information) matrisi olarak adlandırılmaktadır.

ahmin edilen katsayıların varyans ve kovaryans değerleri bu matrisin tersinden elde edilir $\Sigma(\beta) = I^{-1}(\beta)$ şeklinde gösterilir. $\sigma^2(\beta_j)$; bu matrisin j.nci köşegen elemanıdır ve bu β_j 'nin varyansıdır (burada $j = 0, 1, 2, \dots, p$ 'dir). Köşegen elemanları dışındaki elemanlar ise, β_j ve β_u 'nun kovaryans değerleri olup, $\sigma^2(\beta_j, \beta_u)$ ile gösterilmektedir. ($j, u = 0, 1, 2, \dots, p$ 'dir) varyans ve kovaryansların tahmini $\hat{\Sigma}(\hat{\beta}_j)$ ile gösterilmiştir. $\Sigma(\beta)$ değeri ise $\hat{\beta}$ kullanılarak elde edilmiştir. Tahmin edilen katsayıların standart hataları ise şu şekilde hesaplanır;

$$\hat{SE}(\hat{\beta}_j) = \left[\hat{\sigma}^2(\hat{\beta}_j) \right]^{1/2}, \quad j = 0, 1, 2, \dots, p \quad (17)$$

bu formül katsayıların test edilmesi ve güven aralıklarının tahmin edilmesinde kullanılmaktadır. Uyum iyiliği konusunda bilgi (information) matrisinden yararlanılmaktadır.

$$I(\hat{\beta}) = X'VX \quad (18)$$

burada X ve V aşağıdaki gibi ifade edilmektedir;

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

X; $n \times (p+1)$ boyutlu bir matristir olup, herbir birey için verileri içermektedir.

$$V = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

V; $n \times n$ boyutlu bir matris olup, köşegen elemanları $\hat{\pi}_j(1-\hat{\pi}_j)$, ($j = 1, 2, \dots, n$) olan bir köşegen matristir. (1)

2.6.2. Çok deęişkenli modelin önemlilik testi

Çok deęişkenli modeldeki katsayılar tahmin edildikten sonra, modeldeki deęişkenlerin önemli olup olmadıklarının incelenmesi gereklidir. Bu test modeldeki bağımsız deęişkenlerin bağımlı deęişkenle olan ilişkilerinin önemli olup olmadığına testi şeklindedir. Deęişkenlerin önemlilik testinde Wald (W) istatistięi kullanılmaktadır.

Modelin önemlilik testinde ilk adım, modeldeki deęişkenlerin önemli olup olmadığına kontrol edilmesidir. Olabilirlik oran testi, modeldeki bağımsız deęişkenler için p katsayısının tümel önemlilięini verir. Olabilirlik oran testi daha önce anlatılan G istatistięine dayanmaktadır. Modelde bulunan p tane eğim katsayısının sıfıra eşit olması hipotezi altında G istatistięi p serbestlik dereceli ki-kare dağılımına sahiptir. Eğer H_0 red edilirse, en az bir veya bütün p tane katsayının sıfırdan farklı olduęu sonucu elde edilir. Katsayıların hepsinin ya da bazılarının sıfırdan farklı olduęu sonucuna varmadan önce, deęişkenler tek tek Wald test istatistięi ile test edilebilir.

$$W_j = \hat{\beta}_j / SE(\hat{\beta}_j) \quad (19)$$

Bir katsayının sıfıra eşit olması hipotezi altında Wald istatistięi standart normal dağılım göstermektedir. Bu istatistięin önemi, modeldeki herhangi bir deęişkenin önemli olup olmadığını belirtmektedir.

Asıl önemli olan nokta, en iyi uyuma sahip modeli en az parametre ile belirlemektir. Bundan sonraki mantıklı işlem, önemli olduęu düşünölen deęişkenleri modele alıp, deęişken seçim yöntemlerini kullanarak yeni bir analiz yapmak ve modeli bütün deęişkenleri içeren ill model ile karşılaştırmaktır. Burada H_0 hipotezi; “modele alınmayan deęişkenlerin katsayıları 0’a eşittir” şeklinde kurulur. G istatistięi hesaplanır. Burada serbestlik derecesi, iki modelde bulunan deęişken sayılarının farkına eşittir. Bu bilgiler yardımıyla bir karara varılır.

Eđer nominal (kategorik) veya ordinal (sıralanabilir) ölçekli bağımsız deęişkenler modelden çıkarılacak veya eklenecek ise onların dizayn(kukla) deęişkenleri de modelden çıkarılmalı ya da eklenmelidir.(5)

Çok deęişkenli modelde de Wald istatistięi kullanılabilir. Wald istatistięi lanılırken çoklu serbestlik derecesi olduęu durumda dikkat edilmelidir. Wald testinin kullanılabilmesi için vektör-matris işlemlerinin yapılması ve $\hat{\beta}$ vektörünün bulunması gerekmekte ve bu da çok zaman alabilmektedir.

Tek deęişkenli modelin önemlilięinde anlatılan dięer bir test Score testi idi. Score testinin de Wald testi gibi çok deęişkenli modelin önemlilik testinde olabilirlik oran testine göre bir üstünlüęü bulunmamaktadır. Bu nedenle çok deęişkenli modelin önemlilik testinde olabilirlik oran testinin kullanılması tavsiye edilmektedir.(8),(1)

2.7. Lojistik regresyonda sürekli ve kesikli deęişkenlerin kullanıldığı durum

Lojistik regresyon analizinde risk yaklaşımı söz konusu olduğundan sürekli deęişkenlerin kendi özgün deęerleriyle kullanılması, araştırmacıların amacına göre regresyon katsayılarının yorumlanması işleminde sorun olabilmektedir. Risk hesaplamalarında, riskin değerlendirilmesi için riskin hangi gruba göre yapılacağı tespit edilmelidir. Bu nedenle sürekli deęişkenler araştırmacının amacına göre sınıflandırılabilir. Lojistik regresyon analizinde riski değerlendirilmesine göre hesaplanacak sürekli olmayan (nominal veya ordinal) deęişkenin grubuna göre referans (temel) grup adı verilir.

Lojistik regresyon analizinde kullanılacak nominal veya ordinal gibi sürekli olmayan deęişkenler kullanıldığı zaman, deęişkenin düzey sayısı ikiden fazla ise, risk değerlendirilmesinin yapılabilmesi için (düzey sayısı – 1) adet kukla (Dummy) deęişkenle gösterilirler. Bu işlem, regresyon katsayılarının kullanım ve yorumunu kolaylaştırır.

Lojistik regresyonda kukla (dizayn) deęişken oluşturmak için kullanılan yöntemlerden bazıları aşağıdadır. Ordinal deęişkenler için kullanılan bu yöntemlere SPSS for Windows programında yer verilmiştir. SPSS'te seçmeli olarak kullanılan bu prosedürlerde, referans grup ilk ya da sonraki olarak belirlenebilmektedir.(5)

Bu yöntemler:

1. **Tekrarlı (repeated) yöntem** : Her bir grubun riskini, kendinden öncekine göre hesaplanmasına olanak sağlar.
2. **Fark (difference) yöntemi** : Her bir grubun riskini, kendinden öncekilerin ortalama riskine göre hesaplanmasını sağlar.
3. **Helmert Yöntemi**: Her bir grubun riskini, kendinden sonrakilerin ortalama riskine göre hesaplanmasını sağlar.
4. **Sapma (Deviance) Yöntemi** : Her bir grubun riskini, grupların toplam riskine göre hesaplanmasını sağlar.
5. **Basit (Simple) Yöntem** : Her bir grubun riskini, referans (temel) gruba göre hesaplanmasını sağlar.

Tıp ve Sağlık bilimlerinde en çok kullanılan ve tercih edilen yöntem “**basit (simple) yöntemdir**”. Bu yöntemin tercih edilmesinin nedeni, epidemiyolojik ve klinik amaçlara uygun olmasıdır. (18)

2.8. Lojistik regresyon modelindeki katsayıların yorumlanması

Model kurulduktan, tahmin edilen katsayıların hesaplanması ve öneminin değerlendirilmesinden sonra, katsayıların yorumlanması işlemi yapılır. Kurulan bir modelin yorumlanması modeldeki tahmin edilen katsayılar eğimi ya da bağımsız değişkendeki bir birimlik değişimin, bağımlı değişkenin fonksiyonundaki değişim oranını gösterir. Sonuç olarak yorumlama işleminin yapılabilmesi için iki işlem gereklidir:

- (1) Bağımlı ve bağımsız değişkenler arasındaki fonksiyonel ilişkinin belirlenmesi
- (2) Bağımsız değişken için bir birimlik değişmeyi uygun olarak açıklama

Karar verilmesi gereken ilk adım: “ Bağımlı değişkenin hangi fonksiyonu bağımsız değişkenler ile lineer bir fonksiyon oluşturmaktadır?” sorusudur.

Lineer regresyonda bağımlı değişken parametreleriyle lineer olduğundan, link fonksiyonu I (identity) matrisidir. Lojistik regresyonda ise link fonksiyonu lojit dönüşüm olup, diğer link fonksiyonunu kullanan modellere göre bir avantajı, örneklem dizaynının eriye ya da geriye dönük olup olmadığını dikkate almadan etkileri tahmin edebilmektedir.

Lojistik modeldeki etkiler odds'a dayanır ve x 'in bir değeri için tahmin edilen $\pi(x)$ 'in, diğer bir değeri için hesaplanan odds'a bölümü odds ratio (oranı) olarak verilir.

$$g(x) = \ln\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \beta_0 + \beta_1 x_1 \quad (20)$$

$g(x+1) - g(x)$ olup, bağımsız değişkendeki (x 'deki) bir birimlik değişim meydana geldiğinde, lojitte oluşan değişimi ifade eder.

Katsayıların yorumuna bağımsız değişkenlerin ikili olduğu durum ile başlayalım. x 'in 0 ve 1 olarak kodlandığını varsayalım.

blo.2 Bağımlı ve bağımsız değişkenlerin ikili olduğunda lojistik regresyon modelinin parametreleri

		Bağımsız değişken (x)	
		$x = 1$	$x = 0$
Sonuç değişkeni $y = 1$		$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
Sonuç değişkeni $y = 0$		$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Toplam		1.0	1.0

$x=1$ olan bireyler arasında, bağımlı değişkenin görülme ($y=1$) odds değeri $\pi(1)/[1-\pi(1)]$, $x=0$ olan bireyler için bağımlı değişkenin görülme ($y=1$) odds değeri $\pi(0)/[1-\pi(0)]$ olarak verilmiştir. Odds değerlerinin logaritması lojit olarak adlandırılır:

$$g(1) = \ln\left\{\frac{\pi(1)}{1-\pi(1)}\right\}$$

$$g(0) = \ln\left\{\frac{\pi(0)}{1-\pi(0)}\right\}$$

odds oranı ψ ile gösterilip, $x=1$ için odds değerinin, $x=0$ için odds değerine oranı olarak tanımlanmaktadır:

$$\psi = \{\pi(1)/[1-\pi(1)]\} / \{\pi(0)/[1-\pi(0)]\} \quad (21)$$

$\frac{\pi(x)}{1-\pi(x)}$ oranı, lojistik modeldeki birey x 'in riskini ifade etmektedir. $\pi(x)$ ise; tahmin edilen riski temsil etmektedir.

OR'nın logaritması, log- odds, lojit farka eşittir.

$$\ln(\psi) = \ln\left[\frac{\pi(1)}{1-\pi(1)} / \frac{\pi(0)}{1-\pi(0)}\right] = g(1) - g(0)$$

Tablo 2'deki değerler yukarıdaki denklemde yerlerine koyarsak odds oranı aşağıdaki gibi hesaplanır;

$$\psi = \frac{\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \cdot \frac{1}{1 + e^{\beta_0}}}{\frac{e^{\beta_0}}{1 + e^{\beta_0}} \cdot \frac{1}{1 + e^{\beta_0 + \beta_1}}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (22)$$

lojit fark ise; $\ln(\psi) = \beta_1$ olur. Odds oranı, özellikle epidemiyolojide ve biyoistatistikde, $x = 1$ olan birimler arasında bağımlı değişkenin görülmesinin, $x = 0$ olan birimlere göre ne kadar muhtemel (ya da muhtemel değil) olduğunu belirten bir ilişki ölçüsüdür. Bunu bir örnek ile açıklayacak olursak: bağımlı değişkenimiz (y) koroner arter hastalığını, bağımsız değişken olarak (x) diyabet olup olmama durumunu ele alalım. Eğer $\psi = 4$ ise, incelenen bireylerde koroner arter hastalığının, diyabeti olan bireylerde olmayanlara göre 4 kat daha sık görüldüğü ifade edilir.

Sürekli değişkenler için odds oranı yorumlanırken, lojitin sürekli değişkenle arasında lineer bir ilişki olmalıdır. Eğer aralarında lineer bir ilişki yoksa, sürekli değişkenleri gruplara ayırmak veya dizayn (kukla) değişkenler kullanılmalıdır.(8),(1)

Odds oranının elde edilmesini başka bir yol ile anlatalım.

$X_1 = (x_1, x_2, \dots, x_p)$, birinci bireye ilişkin değişkenleri, $X_0 = (x_1, x_2, \dots, x_p)$, ikinci bireye ilişkin değişkenleri ifade etsin. Bu durumda $P(x)$ şöyle olacaktır;

$$P(x) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}} \quad (23)$$

$$OR_{X_1, X_0} = \frac{odds(X_1)}{odds(X_0)} \quad \text{olacaktır.} \quad (24)$$

$$Odds(X_1) = \frac{P(x_1)}{1 - P(x_1)} = e^{-(\alpha + \sum \beta_i x_{1i})} \quad (25)$$

ve

$$Odds(X_0) = \frac{P(x_0)}{1 - P(x_0)} = e^{-(\alpha + \sum \beta_i x_{0i})} \quad (26)$$

şeklindedir.

(25) ve (26) nolu eşitlikler (24) nolu eşitlikte yerine yazılırsa,

$$OR_{X_1, X_0} = \frac{e^{-(\alpha + \sum \beta_i x_{1i})}}{e^{-(\alpha + \sum \beta_i x_{0i})}} = e^{\sum \beta_i (x_{1i} - x_{0i})} = \prod e^{\beta_i (x_{1i} - x_{0i})} \quad \text{olarak bulunur.}$$

Eğer x_i , 0 ve 1 olarak kodlanırsa, $OR_{X_1, X_0} = e^{\beta_i}$ elde edilecektir.(2)

$$P_1(x)/ P_2(x) =0,22 / 0,11 = 2 \text{ elde edilsin.}$$

1. Birey riskinin hesaplanabilmesi için follow – up tipi bir araştırma olmalıdır.
2. Bütün bağımsız değişkenlerin değerlerinin belirlenmesi gereklidir.

2.8.2. Odds ratio (oranı) güven aralığının bulunması

Epidemiyologlar herhangi bir değişkenin katsayısı için güven aralığını bulmaktan çok değişkenin odds oranına ilişkin güven aralığının hesaplanması işlemiyle daha çok ilgilienmektedirler. Bu nedenle burada odds oranına ilişkin güven aralığının hesaplanmasını latacağız.

Odds oranı, β_1 katsayısı için güven aralığının hesaplanması şöyledir:

$$\exp[\hat{\beta}_1 \pm z_{1-\alpha/2} \times SE(\hat{\beta}_1)] \quad (27)$$

Yukarıda odds oranı tahmininin $\psi = e^{\beta_1}$ olduğu belirtilmişti. Bu durum bağımsız değişkenin sadece 0 ve 1 olarak kodlandığında geçerlidir. Genel metot şu şekildedir:

Herhangi bir bağımsız değişkeninin $x = a$ ve $x = b$ değerleri için odds oranı aşağıdaki bidir:

$$\psi(a,b) = \exp[\hat{\beta}_1 x(a-b)] \quad (28)$$

olup, $(a-b) = 1$ olduğunda, $\psi = e^{\beta_1}$ olur. $a = 1$ ve $b=0$ olduğu zaman

$\psi(a,b) = \psi(1,0)$ 'dir.

de edilecek olan odds oranının önemli olup olmadığı araştırılmak istenilsin. Bunu bir örnek e açıklayalım. Bağımlı değişkenimiz KAH olsun. Risk faktörü olarak 0 ve 1 olarak kodlanan gara değişkenini alalım. Sigara değişkeninin katsayısının 0,59 olarak tahmin edilsin.

$H_0 : \beta_s = 0$ olması, OR = 1'e eşit olma durumuyla aynı olacaktır.

$H_1 : \beta_s \neq 0$ veya OR $\neq 1$

Sigara değişkenine ilişkin standart hata değeri ise; 0,35 olsun. $\alpha = 0,05$ yanılma izeyinde güven aralığı hesaplanırsa,

$GA_s = (0,91;3,63)$ olarak hesaplanmış olsun. OR =1 hipotezi altında elde edilen güven alığı 1 değerini kapsadığından H_0 hipotezi red edilemez. Buradan sigara değişkeni için saplanan odds oranının önemli olmadığı ortaya çıkar.

Bağımsız değişken ikili olduğunda dizayn (kukla) değişkenleri oluşturmak için kullanılan iki yöntem: Kısmi metod ve Marjinal metoddur. Örneğin, sigara içme durumu; içmiyor = 1, içiyor = 2 olarak kodlandı ise, kısmi metoda göre kukla (dizayn) değişkeni içmiyor = 0 ve içiyor = 1 olarak kodlanır. Buradaki mantık: kodlanan sayıların en küçüğüne 0, diğerine 1 değeri verilir. Burada kukla değişkeni için tahmin edilen sayının exponansiyeli (üssü) alınır, sigara içenlerin içmeyenlere göre odds ratio'sunun (oranının) tahmini elde edilir.

Tablo 3. Sigara içme durumuna kısmi metodun uygulanması

Sigara içme durumu	Kodlama	Kukla (dizayn) değişkeni (D)
İçmiyor	1	0
İçiyor	2	1

Aynı örneği marjinal metod ile şu şekildedir: kukla değişkeni içmiyor = -1, içiyor = 1 değeri verilir. Buradaki mantık: kodlanan sayıların en küçüğüne -1, diğerine 1 değeri verilir.

Tablo 4. Sigara içme durumuna marjinal metodun uygulanması

Sigara içme durumu	Kodlama	Kukla (dizayn) değişkeni(D)
İçmiyor	1	-1
İçiyor	2	1

Marjinal metod ile sigara içenlerin, içmeyenlere göre odds oranını hesaplamak için yukarıdaki genel formülde a = 1, b = -1 yazılırsa, $\psi = 2\beta_1$ elde edilir.

Yapılan bu kodlama işlemi sadece odds oranını etkilemez, aynı zamanda odds oranı için hesaplanan güven aralığını da etkiler. Marjinal metod kullanıldığında, güven aralığının sınırları aşağıdaki gibidir:

$$\exp[2 \hat{\beta}_1 \pm z_{1-\alpha/2} \times 2 SE(\hat{\beta}_1)]$$

$$\text{Genel formül ise; } \exp[\hat{\beta}_1(a-b) \pm z_{1-\alpha/2}|a-b| \times SE(\hat{\beta}_1)] \quad (29)$$

|a-b| : (a-b)'nin mutlak değeridir.

İkili bir değişken için odds ratio önemli bir parametre olup, lojistik regresyon katsayısı ve odds oranı arasındaki ilişki lojistik regresyon sonuçlarının yorumlanması için temel oluşturur.(8),(1)

2.9. Değişkenlerin etkileşimi (Interaction) ve etki karışımı (counfounding)

Regresyon analizlerinde etkileşim (interaction) ve etki karışımı (counfounding) modelin oluşturulmasında önemli olan unsurlardır. Etkileşim ve etki karışımı incelenirken, kullanılan bağımsız değişkenlere ek olarak yardımcı (kontrol) değişkenleri de kullanılabilir.

Etkileşim; Kontrol değişkeninin farklı düzeylerinde bağımsız değişkenle bağımlı değişken arasındaki ilişkinin farklı olmasıdır. Bu durumda kontrol değişkeni, bağımsız değişkenin bağımlı değişken üzerindeki etkisinde değişikliğe sebep olur. Bu durumu neden olan değişkene etki değiştirici denilmektedir. Etkileşimin anlamlı bulunması durumunda bağımsız değişkene ilişkin risk oranları kontrol değişkeninin her bir düzeyi için ayrı ayrı değerlendirilip, yorumlanmalıdır. Etki değiştirici bir kontrol değişkeninin etkisi, tabakalı çalışma yapılarak kontrol edilebilir. (5)

Etki karışımı; kavram olarak etkileşimle karışmaktadır; ancak, etkileşim ve etki karışımı aynı değildir. Etki karışımını olup olmadığını belirlemek etkileşimde olduğu gibi kolay değildir. Etkileşimde kontrol değişkeninin her bir düzeyinde, bağımlı- bağımsız değişken ilişkisi farklı iken, etki karışımında bu durum söz konusu değildir. Kontrol değişkeninin etkisi bağımsız – bağımlı değişken ilişkisi içinde gizlidir. Bu gizlilik kontrol değişkeninin hem bağımlı hem de bağımsız değişkenle ilişkisinden kaynaklanmaktadır. Bu nedenle bağımsız değişkenle, bağımlı değişken ilişkisinin etkileşimde olduğu gibi kontrol değişkeninin her bir düzeyine göre incelenmesi, kontrol değişkeninin etkisini gidermez. Etki karışımını gidermek için kontrol değişkeninin etkisini ortadan kaldırmak gerekir. (18)

Etki karışımı, epidemiyolojistler tarafından hem sonuç değişkeni hem de primer bağımsız değişken ya da risk faktörü ile ilişkili olan bir birlikte değişeni tanımlamak için kullanılmışlardır. İkili bir risk faktörü değişkeni ve sürekli bir birlikte değişene sahip olan bir modelimiz bulunsun. Eğer birlikte değişen ve sonuç değişkeni arasındaki ilişki risk faktörünün her bir düzeyi için aynı ise, o zaman birlikte değişen ve risk faktörü arasında etkileşim yoktur. Grafikselsel olarak etkileşimin olmaması, birbirine paralel olan iki çizgili bir model verir, bu çizgilerin her biri risk faktörünün bir düzeyi içindir. Genel olarak, etkileşimin olmaması model tarafından iki veya daha fazla değişkenden oluşan iki veya daha yüksek derecede terim olmamasıyla kendini gösterir.

Etkileşim olduğu zaman, risk faktörü ve sonuç değişkeni arasındaki ilişki farklı olur ya da birlikte değişenin düzeyine bazı yönlerden bağımlı olur. Yani birlikte değişen, risk faktörünün etkisini değiştirir.

Etki karıştırıcı statüsüne karar vermek için iki koşul gerekir:

1. Birlikte değişenin sonuç değişkeniyle ilişki içinde olması.
2. Birlikte değişenin risk faktörüyle ilişki içinde olması.

Uygulamada, birlikte değişenin risk faktörü olup olmadığı, birlikte değişeni kapsayan ve kapsamayan modellerden elde edilen risk faktörü için tahmin edilen katsayıların karşılaştırılmasıyla anlaşılır. Risk faktörü için tahmin edilen katsayıdaki “biyolojik olarak önemli” herhangi bir değişim, birlikte değişeni etki karıştırıcı olarak belirler ve birlikte değişen için tahmin edilen katsayının istatistiksel önemine bakılmadan modele dahil edilmelidir. Diğer yandan, bir birlikte değişenin etki değiştirici olduğu düşünülüyorsa, etkileşim teriminin modele eklenmesi hem istatistiksel hem de biyolojik olarak önemlidir. (1),(8)

2.10. Değişken Seçimi

Eğer araştırmacının amacı, iyi bir tahmin modeli elde etmek ise, yani tahmin ise, bu durumda bilgisayar algoritmaları (geriye doğru eleme, ileriye doğru seçim,...) gibi bilgisayar tarafından yapılan yöntemlerin kullanılması daha uygun olacaktır.

Lineer regresyonda değişkenlerin seçim işleminde adımsal seçim sıkça kullanılmaktadır. Adımsal lojistik regresyonun metodolojisi kısa zamanda mümkün olmaktadır. Adımsal seçim yöntemini uygulamak, çok sayıda değişkenin hızlı ve etkili bir biçimde incelenmesini ve bazı lojistik regresyon denklemleri ile eş zamanlı uyum sağlar.

Modelden değişkenlerin çıkarılması ya da seçimi için herhangi bir adımsal yöntem, değişkenin önemini kontrol eden ve belli kurallara göre değişkenleri modele dahil eden ya da modelden çıkaran istatistiksel bir algoritmayı temel alır. Bir değişkenin önemi, değişkenin katsayısının istatistiksel anlamlılığını ölçme açısından tanımlanır. Kullanılan istatistik, modelin varsayımlarına dayanır. Hataların normal dağıldığı varsayıldığına göre adımsal doğrusal regresyonda F testi kullanılır. Lojistik regresyonda hatalar binom dağılımına sahip olup, anlamlılığı ise ki-kare testi olasılık oranına göre belirlenir.

Böylece istatistiki anlamda, en önemli değişken yöntemin herhangi bir adımında değişkeni kapsamayan bir modele kıyasla logaritmik olasılıklardaki en fazla değişikliğe neden olan değişkendir. (Yani en fazla istatistik olabilirlik oran istatistiğine (G) sahip değişkendir.)

k düzeyli bir değişken, k-1 tane kukla (dizayn) değişkeni kullanılarak bir model oluşturulsun. G büyüklüğü serbestlik derecesine bağlı olduğundan, G'yi (olasılık oranı test istatistiği) temel alan herhangi bir yöntem, değişkenler arasındaki olabilecek serbestlik derecesi farklılıklarını dikkate alınmalıdır.(6)

Adımsal lojistik regresyonda, öncelikle ileriye doğru seçim yöntemi anlatılıp, sonra geriye doğru eleme yöntemi anlatılacaktır. İşlemin her bir adımında bilgisayar tarafından yapılması gereken hesaplamalar dikkate alınarak açıklanacaktır.

Adım 0 : p tane bağımsız değişkenin olduğunu kabul edelim. Bu değişkenlerin hepsi bağımlı değişkeni açıklarken biyolojik olarak önemli oldukları varsayalım. Adım 0'da, sadece sabitin bulunduğu model kurulup, kurulan modelin log-olabilirliğinin (L_0) hesaplanmasıyla başlar. P değişkenin herbiri için lojistik regresyon modeli kurulduktan sonra, onların log-olabilirlikleri sırasıyla L_0 ile karşılaştırılır. Adım 0'da x_j değişkenini kapsayan modelin log-olabilirlik değeri $L_j^{(0)}$ ile gösterilsin. j; modele dahil edilen değişkeni temsil ederken, 0 üst yazısı adım sayısını gösterir. Bu gösterimler adımsal lojistik regresyon konusu boyunca adım sayısını ve modeldeki değişkenleri takip edebilmek için kullanılacaktır. x_j değişkenini kapsayan modelin yalnızca kesim noktasını kapsayan modele karşı olabilirlik oran test değeri $G_j^{(0)} = 2(L_j^{(0)} - L_0)$ ile gösterilip, p değeri $P_j^{(0)}$ ile gösterilmiştir. $P[\chi^2(v) > G_j^{(0)}] = P_j^{(0)}$ ile p değeri belirlenir. Eğer x_j sürekli bir değişken ise, $v = 1$ ve eğer x_j kesikli ve k düzeyli olan bir değişken ise, $v = k-1$ olarak kabul edilir.

En küçük p değerine sahip olan değişken en önemli değişkendir. Eğer bu değişken x_{e_1} ile gösterilirse, $p_{e_1} = \min(P_j^{(0)})$ eşittir. "e₁" alt yazısı, adım 1'de modele girmeye aday değişkeni göstermek için kullanılır. Eğer x_2 en küçük p değerine sahip ise, bu durumda $P_2^{(0)} = \min(P_j^{(0)})$ eşittir ve $e_1=2$ olur. x_{e_1} en önemli değişken olsa da istatistiksel olarak da önemli olmasını gerektirmez.

Değişkenlerin önemliliklerine karar verirken kullanılacak olan alfa (α) düzeyinin seçimi adımsal lojistik regresyonda önemlidir. P_E herhangi bir değişkenin modele girebilmesi için kullanılacak alfa düzeyi olsun. P_E değerinin küçük veya büyük olarak seçimi modele girecek olan değişken sayısını etkileyecektir. Adımsal lineer regresyonda P_E 'nin seçimi üzerinde Bendel ve Afafi (1977) tarafından çalışmalar yapılmış ve adımsal diskriminant analizi için bu konu üzerinde Costanza ve Afafi durmuşlardır. Araştırmaları sonucunda $P_E = 0,05$ olarak seçilmesi çoğu zaman önemli değişkenleri model dışında bırakabildiği görülüp, P_E 'nin 0,15 ile 0,20 arasında seçilmesi önerilmekte ve adımsal lojistik regresyon analizinde de bu değerlerin kullanılması iyi sonuç vermektedir.

Bazı araştırmacılar çalışmalarının amacını geniş tutup, çok sayıda değişkeni içeren modeli oluşturmak isteyebilirler. Böyle durumlarda P_E 'nin 0,25 olarak alınması daha iyi sonuç verebilir.

P_E 'nin değeri her ne olursa olsun, G için p değeri P_E 'den küçükse, o zaman değişken modele dahil edilebilecek kadar önemli olduğu sonucuna ulaşılır. Eğer $P_{e_1}^{(0)} < P_E$ ise, Adım 1'e geçilir, yoksa işlem durdurulur.(8)

Adım 1 : Adım 1, x_{e_1} değişkenini içeren modelin kurulmasıyla başlar. Modelin log-olabilirliği $L_{e_1}^{(0)}$ ile gösterilsin. x_{e_1} modele girdikten sonra, geriye kalan $p-1$ değişkenin önemliliğini kontrol etmek için x_{e_1} ve x_j 'yi ($j=1,2,\dots,p$ ve $j \neq e_1$) kapsayan $p-1$ tane lojistik regresyon modeli kurulur. x_{e_1} ve x_j 'yi ($j=1,2,\dots,p$ ve $j \neq e_1$) kapsayan modelin log-olabilirlik değeri $L_{e_1,j}^{(0)}$ ile ifade edilsin. Olabilirlik oran test istatistiği $G_j^{(1)} = 2 (L_{e_1,j}^{(1)} - L_{e_1}^{(0)})$ olarak verilmiştir. G istatistiği için p değeri $P_j^{(1)}$ ile ifade edilip, en küçük p değerine sahip değişken x_{e_2} olsun ($P_{e_2}^{(1)} = \min (P_j^{(1)})$). Bu değer P_E değerinden küçükse Adım 2'ye geçilir, P_E değerinden büyükse işlem durdurulur.

Adım 2 : Adım 2, x_{e_1} ve x_{e_2} değişkenlerini içeren modelin kurulmasıyla başlar. x_{e_2} modele girdikten sonra x_{e_1} değişkeninin önemli olup olmadığı kontrol edilmelidir. Bu işlem ile Adım 2 geriye doğru eleme kontrolü işlemini de kapsamaktadır. Genel olarak bu işlem, bir önceki adımda eklenen değişkenlerden birisi silinerek model kurulur ve ardından silinen değişkenin öneminin kontrol edilmesi işlemiyle gerçekleştirilir. Adım 2’de, x_{e_1} değişkeni çıkarıldıktan sonraki modelin log-olabilirliği $L_{-e_1}^{(2)}$ ile gösterilsin. Adım 2 de Modelin full modele karşı olabilirlik oran testi $G_{-e_1}^{(2)} = 2(L_{e_1 e_2}^{(2)} - L_{-e_1}^{(2)})$ ve p değeri de $P_{-e_1}^{(2)}$ olarak gösterilir. Modelden bir değişkenin çıkarılıp çıkarılmayacağından emin olmak için program, değişken çıkarıldığı zaman en büyük p değerine sahip olan değişkeni seçer. Bu değişken x_{r_2} olarak verilir, $p_{r_2}^{(2)} = \max(p_{-e_1}^{(2)}, p_{-e_2}^{(2)})$. x_{r_2} değişkeninin modelden çıkarılıp çıkarılmayacağına karar verebilmek için, program $p_{r_2}^{(2)}$ değerini, önceden seçilmiş olan ikinci bir “alfa” düzeyiyle (P_R) karşılaştırılır. P_R değeri modele katkıyı devam ettirebilmek için minimal düzeyi gösterir ve R alt yazısı modelden çıkartılmayı temsil eder. Programın birbirini takip eden adımlarında aynı değişkenin modele dahil edilmesi ve modelden çıkarılması gibi bir olasılığı engellemek için, P_R değeri ne olursa olsun, P_E değerinden büyük olmalıdır.

Eğer modelden çıkartmak için maksimum p değeri ($p_{r_2}^{(2)}$) P_R değerini geçerse, x_{r_2} modelden çıkarılır ($p_{r_2}^{(2)} > P_R$). Eğer $p_{r_2}^{(2)} < P_R$ ise, x_{r_2} modelde kalır.

İleriye yönelik seçim safhasında x_{e_1} , x_{e_2} ve x_j 'yi ($j=1,2,\dots,p$ ve $j \neq e_1, e_2$) içeren p-2 tane lojistik regresyon modeli kurulup, herbir modelin log-olabilirlik değeri hesaplandıktan sonra, sadece x_{e_1} ve x_{e_2} 'yi kapsayan modele karşı olabilirlik oran testlerini hesaplanır ve karşılık gelen p değerleri bulunur. x_{e_3} en küçük p değerine sahip olan değişken olsun ($P_{e_3}^{(2)} = \min(P_j^{(2)})$). Eğer $P_{e_3}^{(2)} < P_E$ ise, Adım 3’e geçilir, aksi halde işlem durdurulur.

Adım 3 : Bu adımda da bir önceki adımda (Adım 2’de) yapılan işlemler ile aynıdır. Program önce geriye doğru elemeyi kontrol ettikten sonra ileriye doğru seçim yapar. Bundan sonraki işlemler aynı mantıkla program tarafından en son adıma (Adım S) kadar devam eder.

Adım S : Bu adımda iki muhtemel durum ortaya çıkar:

1. Değişkenlerin hepsinin (p değişkenin) modele girmesi.
2. Modeldeki bütün değişkenlerin bütün değişkenlerin model dışı kalması için geçerli p değerleri, P_R değerinden küçük olduğu zaman ve modele alınmayan değişkenlerin modele girebilmeleri için gerekli p değerleri, P_E değerinden büyük olduğunda Adım S'e geçilir. Bu adımda P_R ve P_E kriterlerine göre önemli olan değişkenler modelde bulunurlar. Eğer P_R ve P_E 'nin değerleri istatistiksel önemliliğe ulaşmak için güvenilir değerlerde seçilirse, o zaman final model için değişkenler, adimsal regresyon işlemini özetleyen bir tablodan seçilecektir.

Özet tablodan değişken seçimi için kullanılabilir iki yöntem vardır:

1. Her bir adımda modele alınma kriteri p değerine dayanmaktadır.
2. Şu an ki adımdaki modele karşı son adımdaki modelin olabilirlik oran testine dayanmaktadır.

Mesela herhangi bir adımı q ile ifade edelim. Birinci yöntemde $P_{e_q}^{(q-1)}$ değeri, önceden belirlenmiş olan anlamlılık düzeyi ile karşılaştırılır. Eğer $P_{e_q}^{(q-1)} < \alpha$ ise Adım q'ya geçilir. $P_{e_q}^{(q-1)} > \alpha$ ise, işlem bitirilir. Bu yöntemde modele girme kriteri, $\chi_{e_1}, \chi_{e_2}, \dots, \chi_{e_{q-1}}$ değişkenlerinin modelde bulunması şartıyla χ_{e_q} 'nin katsayısının önem testine dayanmaktadır. Testin serbestlik derecesi, χ_{e_q} 'nin k kategorili veya sürekli olması durumunda sırasıyla k-1 veya 1'dir.

İkinci yöntemde, şu an ki adımda ki (Adım q) modeli, bir önceki adımdaki (Adım q-1) modelle değil de, en son adımdaki (Adım S) modelle karşılaştırılır. Bu iki modelin olabilirlik oran testi için p değeri bulunur ve $p > \alpha$ oluncaya kadar işlem devam eder.

Bu olay Adım q'dan Adım S'ye kadar modele eklenen değişkenlerin hepsinin katsayılarının sıfıra eşit olmasını test eder. İkinci yöntem kullanıldığında serbestlik derecesi birinci yöntemde kullanılan serbestlik derecesinden daha büyük olduğundan, ikinci yöntem, birinci yöntemden daha çok sayıda değişken seçebilir.

Adımsal seçim işleminde hesaplanan p değerlerinin geleneksel hipotez testlerinden farklı bir değerdir. Bu p değerlerinin değişkenler arasındaki göreceli önemin bir göstergesi oldukları düşünülür. Göreceli olarak zengin bir model isteniliyorsa adımsal seçim yöntemi tavsiye edilebilir.

Adımsal yöntemin en belirgin özelliği, işleme önemli kovaryansları kapsayan Adım 0 ile başlamasıdır. Seçim işlemi diğer değişkenlerle devam eder.

Adımsal seçim işlemi, her bir adımda modelde olmayan değişkenlerin de katsayılarının maksimum olabilirlik tahminini hesaplamakta, bu da çok zaman almakta ve maliyeti arttırmaktadır.

Adımsal seçim yöntemleri, kurulan model için sadece istatistiksel metodlara dayanarak aday değişkenleri belirler. (1)

2.11. Çoklu Regresyonda Değişken Seçimi

Çoklu regresyon analizinde verileri en iyi şekilde temsil edecek değişkenler seçmek, yani bağımlı değişkeni en iyi açıklayan değişkenleri seçmek ve modele çok az katkısı olan değişkenleri modelden çıkarmaya değişken seçme adı verilir. Mesela Koroner Arter hastalığı bağımlı değişken olarak alınırsa, KAH'nı en çok açıklayan değişkenler modele alınırken, KAH'nı açıklamaya katkısı olmayan değişkenler ise modelden çıkarılacaktır.

Regresyon analizinde, artık değişkenlerin ve çoklu bağlantıların incelenmesinden sonra, en iyi model oluşturulurken önemli değişkenler seçilmelidir. Modelin kurulmasıyla beraber, çoklu bağıntıları çözümlmek için de bu yöntem kullanılmaktadır.

Regresyon analizinde bütün değişkenlerin bulunduğu model için H_0 hipotezi test edilir. H_0 red edilirse, "en az bir j değeri 0'dan farklıdır" şeklinde yorum yapılır. Böyle bir durumda hangi j'nin sıfırdan farklı olduğu araştırılmalıdır. Kısmi F değerlerinin F_α 'dan küçük olduğu değişkenler modele katkısı olmayan değişkenlerdir. Bu durum (bütün kısmi F'lerin incelenmesi) sadece bağımsız değişkenlerin birbirlerine dik olduğu durumda geçerlidir. Bağımsız değişkenler arasında bir etkileşimin olması bunu geçersiz yapmaktadır.

Bütün kümeler $\beta_1, \beta_2, \dots, \beta_k = 0$ kısmi F'den kabul edilebilir. Ancak x_1 değişkeni modelden çıkartılıp k-1 değişken üzerinden inceleme yapıldığında, $\beta_2 = 0$ bu kez red edilebilir.

Değişken seçiminin yapılmasının başlıca nedenleri şöyle sıralanabilir:

1. Herhangi bir modelde az sayıda değişken bulunmasının uygulama ve maliyet bakımından yararı vardır. Modelde bulunan değişkenleri sonradan izlemek gerekebileceğinden bu değişkenlere ilişkin verileri toplamada ortaya çıkabilecek güçlükler veya verileri elde etme maliyetinin yüksek olması araştırmacıları az sayıda değişkenle araştırmalarını yapmaya sevk eder.
2. Tahminlerin ve ön tahminlerin istenilen istatistiki özelliklere sahip olmaları gerekir.(6)

2.12. Adımsal Teknikler ve Durdurma kuralları

Adımsal tekniklerde durdurma kuralı olarak, çoğu kez F kullanılır. Bu değerler bir değişkenin artık kareler toplamında meydana getirdiği küçülmeyi ölçer. Modele bir değişken eklendiği zaman artık kareler toplamı azalır veya bunun tersine modelden bir değişken çıkartıldığında artık kareler toplamında bir artma olur. Değişken ne kadar önemli ise bu artma miktarı da o kadar büyük olur. Böylece bu kısmi F değerleri, bir değişken denkleme eklenmiş veya denklemden çıkartılmış gibi düşünülerek, bu değişkenin artık kareler toplamında yaptığı değişimi test etmeye yardımcı olur.

2.12.1. İleriye doğru değişken seçim yöntemi

Bu yöntemin mantığı, regresyon denkleminde hiç bir değişken bulunmazken, değişkenlerin her adımda denkleme birer birer eklenmesi temeline dayanır. İlk adımda bağımlı değişkeni en fazla açıklayan başka bir ifadeyle bağımlı değişkenle en yüksek ilişkisi olan değişken denkleme alınır. Model ile ilgili F testi yapılır. Eğer test sonucu önemli ise, ikinci ve daha sonraki adımlarda bağımlı değişkenle en çok kısmi ilişkiye sahip olan değişken denkleme alınır. Modele en son dahil edilen değişken β_i olsun. β_i için H_0 hipotezi $H_0 : \beta_i = 0$ hipotezi kısmi F testi ile test edilir.

Kısmi F testi formülü aşağıdaki gibidir:

$$F_{(i)} = \frac{AKT_p - AKT_{p+(i)}}{\sigma^2_{p+(i)}} \quad (30)$$

AKT_p ; bir önceki modelden elde edilen artık kareler toplamıdır.

$AKT_{p+(i)}$; i.nci değişken modele dahil edildikten sonra $(p+i)$ sayıdaki değişken üzerinden bulunan artık kareler toplamıdır. Böylece $AKT_p - AKT_{p+(i)}$ denklemi i.nci değişkenin eklenmesiyle Artık kareler toplamında elde edilen azalmayı verir. Eğer herhangi bir aşamada F_i değeri F tablo değerinden büyük ($F_i > F_{\text{Tablo}}$) ise $H_0 : \beta_i = 0$ hipotezi red edilir. H_0 hipotezinin red edilmesi ile β_i önemli olup, i.nci değişken modele eklenmiş olur. $F_i > F_{\text{Tablo}}$ ise, işlem bitirilir. Bir önceki modelin en iyi alt küme denklemi olduğu sonucuna varılır.

2.12.2. Geriye Doğru Değişken Çıkarma Yöntemi

İlk olarak mevcut bütün değişkenlerin bulunduğu regresyon denkleminde değişkenlerin her adımda tek tek çıkarılmasıyla uygulanmaktadır. Herhangi bir adımda denklemdaki her bir değişken için kısmi F değerleri hesaplanır.

$$F_{(i)} = \frac{AKT_{p-(i)} - AKT_p}{\sigma^2_p} \quad ; (i= 1,2,\dots,p \text{ için}) \quad (31)$$

ile hesaplanır.

AKT_p : p değişkenli modelden elde edilen artık kareler toplamıdır.

$AKT_{p-(i)}$: i.nci değişken modelden çıkartılarak elde edilen artık kareler toplamıdır. Her adımda p değişken için artık kareler toplamı hesaplanır. Eğer en küçük $F_{(i)}$, F tablo değerinden küçük ise, i.nci değişken denklemden çıkartılır. $F_{(i)} > F$ tablo değeri ise işlem bitirilir. Böylece en iyi model denkleminin i.nci değişkenin de bulunduğu alt küme denkleminin olduğu sonucuna varılır. (6)

2.13. Uyum İyiliği

Doğrusal regresyon çalışmalarında, modelin önemliliği için yapılan varyans analizi gibi analizler, lojistik regresyonda da modelin uyum iyiliği için gereklidir. Uyum iyiliğinde kullanılan istatistik ve analizler çeşitlidir. Bunlardan bir kısmı regresyondan ayrılışın test edilmesi gibi klasik anlayışı taşırlar. Diğerleri ise lojistik mantıktan hareket ederek χ^2 (Ki-kare) uyum iyiliği istatistiğini kullanmaktadırlar.

Lojistik varsayım çok değişkenli dağılımların geniş bir kesimi için sağlansa da kurulan lojistik modelin geçerliliği kontrol edilmelidir. Modelde bulunması gereken bütün değişkenler ele alındıktan sonra modelin cevap değişkenini açıklamadaki etkinliğini araştırmaya “ Uyum iyiliği” denilmektedir. Lojistik modelin uyum iyiliğini araştırmada kullanılabilecek ölçütler:

1. Bütün değerleri içeren model ile tahmin edilen modele ilişkin olabilirlik oran değerlerinin farkına dayanan artık kareler toplamına (AKT) benzer ölçütler χ^2 dağılımlı olup modelin geçerliliğini sınamada kullanılırlar. Bu yol ile modele eklenen karesel terimin etkisi de sınanabilmektedir.

2. Artık değerler hesaplanarak bunların x'e ya da olasılık değerlerine karşı çiziminden aykırı değerler araştırılır. Bulunacak olan bu aykırı değer, model uyumundaki sorunun göstergesi olacaktır.

3. AKT ve olabilirlik oranına dayalı R^2 gibi ölçütler vardır. Uyum testi için yapay (pseudo) R^2 ölçütleri incelenebilmektedir.

4. Lojistik model, ayırmsama amacı ile kullanıldığında modelin doğru sınıflandırma yüzdeleri de birer uyum iyiliği ölçütü olarak kullanılabilir. Bu yüzdeler yararlı bir ölçüt olmasına rağmen ayırmsama gücünün yeterli bir göstergesi değildir.

Lojistik modelde normallik varsayımı sağlanmadığı için modelin uyum iyiliği testlerinde t ve F istatistiklerinin yerine χ^2 ve D gibi parametrik olmayan ölçütler kullanılmalıdır.(5)

2.13.1. D (Deviance) ve Pearson χ^2 (Ki-kare) Ölçütleri

Regresyondan ayrılışı test etme amacını taşıyan uyum iyiliği test istatistiklerinde bağımsız değişken kavramı kullanılmaktadır. Çok değişkenli regresyon analizlerinde birbiri ile aynı olan bağımsız değişken kombinasyonları bulunmaktadır. Bu kombinasyonlara bağımsız değişken deseni (covariate pattern) denilmektedir. m ; bağımsız değişken deseni sayısını ifade etsin. Bağımsız değişken deseni sayısı (m) \leq toplam gözlem sayısıdır (n). Bağımsız değişken desenleri regresyonda aynı davranışı göstermediklerinden, özellikle regresyondan ayrılışların hesaplanmasında önemli rol oynarlar. Regresyondan ayrılışlar;

$$(Artık) A_i = y - P(y=1, x), \quad (i = 1, 2, \dots, m)$$

i .nci değişken deseni için;

Z_{A_i} ; Standartlaştırılmış artık olarak ifade edilmektedir. Bu durumda Pearson (χ^2) istatistiği;

$$\text{Pearson } \chi^2 = \sum_{i=1}^m Z_{A_i}^2 \quad (32)$$

eşitliği ile bulunur. Pearson istatistiği ki-kare (χ^2) dağılımına sahip olup, serbestlik derecesi, k bağımsız değişken sayısını göstermek üzere ($m-k-1$) olur.

Regresyon ayrılışın incelenmesinde kullanılan diğer bir istatistik de Deviance istatistiğidir. (x_i, y_i) çiftinin gözlem sayısı 1 olmak üzere Deviance artıkları şöyledir;

$$y_i = 1 \quad \text{için} \quad d_i = -\sqrt{2|\ln(P(y=1, x))|} \quad (33)$$

$$y_i = 0 \quad \text{için} \quad d_i = -\sqrt{2|\ln(1 - P(y=1, x))|} \quad (34)$$

şeklinde ifade edilirler. Bu özel durumun dışında i .nci değişken desenindeki gözlem sayısı n_i olmak üzere,

$$d_i = \left\{ 2 \left[y_j \ln \left(\frac{y_i}{n_i P(y_i, x_i)} \right) + (n_i - y_j) \ln \left(\frac{(n_i - y_j)}{n_i [1 - P(y_i, x_i)]} \right) \right] \right\}^{1/2} \quad (35)$$

ve

$$D = \sum_{j=1}^m d_j^2 \quad (36)$$

eşitlikleriyle gösterilir. D istatistiği de (m-k-1) serbestlik derecesiyle ki-kare dağılımı gösterir. Pearson ve Deviance istatistiklerinin m = n olduğu durumlarda dağılımları bozulduğundan kullanılmaları tavsiye edilmemektedir.(12),(17)

2.13.2. Diğer uyum iyiliği ölçütleri

Doğrusal regresyon analizinde modeldeki katsayıların anlamlılığı F testi ile belirlenmektedir. Lojistik regresyon analizinde bu teste karşılık olarak (6) eşitliği ile tanımlanan sapma ölçütü kullanılabilir. Payda ki tek bir sabit terim içeren modelin olabilirlik değerinin (L_0), paydada ise elde edilen modelin olabilirlik değerinin (L_1) yer aldığı ölçüt,

$$C = -2 \ln(L_0/L_1) = -2 (\ln L_0 - \ln L_1) \quad (37)$$

Şeklinde olup p-1 serbestlik dereceli ki-kare dağılımı göstermektedir.

Öte yandan, C'ye benzer olarak en iyi modele ulaşmak için çeşitli modellerin olabilirlik değerleri birbirleriyle (6) nolu sapma ölçütü yardımıyla kıyaslanabilmektedir. Modellerin içerdikleri değişken sayıları arasındaki fark, ki-kare dağılımlı ölçütün serbestlik derecesini oluşturmaktadır.

Lojistik modelin uyum iyiliğinde kullanılacak en basit ölçülerden biri, doğrusal regresyonda modelin açıklanma miktarı R^2 'ye karşılık gelen yapay (pseudo) R^2 olup,

$$\text{Pseudo- } R^2 = C/(C+n) \quad (38)$$

Burada C , (37) nolu formülde verildiği gibidir. 0-1 arasında değerler alan bu ölçütün sıfıra yakın değeri modelin uyum eksikliğini bir göstergesi olmaktadır.

Bunlara ilaveten diğer bir ölçüt, modelin log-olabilirlik değerinin -2 katı şeklinde olup T1 ölçütü adını almakta ve (n-p-1) serbestlik dereceli ki-kare dağılımı göstermektedir.

Hilden ve arkadaşları elde edilen çeşitli modeller arasında seçim yapmak için modellerin doğru sınıflandırma olasılıklarına dayanan T7 ölçütünü kullanmışlardır. Bu ölçüte göre en büyük T7 değerine sahip olan model en iyi model olarak tanımlanmıştır.

Uyum eksikliğini verileri gruplara bölerek test edebilen ölçütler de bulunmaktadır. Bunlardan birisi, aralıklara bölünerek kesikli hale dönüştürülen sürekli bağımsız değişkene ait bölünmeden önce ve sonraki olasılık fonksiyonlarının farkı olarak ifade edilmektedir. Ki-kare dağılımına sahip ölçüt, Landwehr ve arkadaşlarının (1984) önerdiği lokal- ortalama sapma (local mean deviance) adlı grafiksel yaklaşımın özel bir durumudur. Bu grafiksel yaklaşımda ise veri kümesi gruplara ayrılarak, her grubun sapma ölçütüne lokal etkisi araştırılmaktadır. Bu etkiler global etki ile karşılaştırılarak modelin uyum eksikliğinin olup olmadığına karar verilmektedir.

Tsiatis (1980) tarafından bir uyum eksikliği ölçütü önerilmiştir. Bağımsız değişkenleri gruplara bölerek hesaplanan ve (grup sayısı -1) serbestlik dereceli ki-kare dağılımına sahip olan ölçüt, lokal ortalama sapma çizimlerine karşılık gelmekte ve skor test istatistiği adını almaktadır. Skor test istatistiği;

$$ST = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sqrt{\bar{y} (1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (39)$$

şeklinde ifade edilmektedir. Normal dağılıma sahip skor test istatistiği, α anlamlılık düzeyinde kritik z değeri ile karşılaştırılarak karar verilir.

Hosmer-Lemeshow, bağımsız (açıklayıcı) değişkenlerin bölünmesi işlemini olasılık tahminlerine dayanarak yapan Hosmer-Lemeshow $HL(\hat{C})$ test ölçütünü önermişlerdir. Ölçüt, veriler g gruba bölündüğünde 2xg çapraz tablosundan Pearson χ^2 değerlerinin hesaplanmasına dayanmakta olup, (g-2) serbestlik dereceli ki-kare (χ^2) dağılımına sahiptir. Kolay yorumlanması bir avantajı iken verileri gruplandırmadan kaynaklanan bilgi kaybının uyumdan sapmaları göz ardı etmeye neden olabilmesi de dezavantajıdır.

HL test ölçütü aşağıdaki gibi ifade edilmektedir;

$$HL(\hat{C}) = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (40)$$

Modelin sınıflandırma tablosundan elde edilen doğru- yanlış sınıflandırma oranları da bir uyum iyiliği ölçütü olup, analizin amacı sınıflama olduğu zaman uygun bir ölçüttür; yoksa uyuma karar vermek için kullanılan yöntemlere ek bir katkısı olabilir. Tabloda bağımlı değişkenin gerçek ve tahmin değerleri çaprazlanmaktadır. Tahmin değeri, 0,5'den büyükse 1, 0,5'den küçükse 0 grubuna yerleştirilmektedir .Doğru-yanlış sınıflandırma oranının uyum iyiliği kriteri olarak kullanmanın bazı sakıncaları vardır. Sürekli olarak tahmin edilen cevap değişkeni kritik değerlerin yardımıyla kesikli hale getirilebilir. Bu durumda 0,5'e çok yakın olan değerlerin farklı gruplara atanması söz konusu olabilmektedir. Bir örnek ile açıklayacak olursak; $P = 0,49$ değeri ile $P = 0,51$ değeri arasında 0,02'lik bir fark vardır. 0,5 kritik değeri ile karşılaştırıldığında çok yakın olan bu iki değer farklı gruplara atanacaktır.(5),(6)

2.14. SPSS'de çoklu regresyon analizi yöntemleri

Bir araştırmacı bağımlı ve bağımsız değişkenlerini belirledikten sonra, SPSS 8.0 for Windows paket programını kullanacaksa, analiz yapmak için şu yol izlenir;

SPSS / Statistics / Regression / Logistic yolu izlenmelidir. SPSS'de bulunan regresyon yöntemleri şunlardır;

1. **Enter Yöntemi:** Bütün bağımsız değişkenlerin tek adımda değerlendirildiği yöntem Enter yöntemi denilmektedir.

2. **İleriye Doğru Seçim (Forward Selection) Yöntemi:** Bağımlı değişken ile en yüksek (pozitif veya negatif) korelasyona sahip olan bağımsız değişken ilk olarak modele alınır. Modele alınan değişkenin katsayısı β_i olsun. $H_0 : \beta_i = 0$ hipotezi altında, H_0 hipotezi F testi ile test edilir.

Hesaplanan F değeri, SPSS'in F değeriyle karşılaştırılır. SPSS'de F değeri iki şekilde tespit edilmektedir;

(a) Kullanıcı tarafından belirlenen minimum bir F değeri.

(b) F istatistiği ile ilişkili olan bir ihtimalin belirlenmesi (PIN; Probability of F-to-enter). Bunun normal değeri 0.05'dir.

Hesaplanan F değeri SPSS'in F değerine eşit veya küçükse, ilgili bağımsız değişken modele alınır ve seçim işlemi devam eder. Aksi halde işlem durdurulur.

Bir bağımsız değişken modele alındıktan sonra, geriye kalan bağımsız değişkenlerle bağımlı değişken arasındaki korelasyona bakılır. En yüksek korelasyona sahip olan bağımsız değişken modele girmeye aday olur. Bu, aynı zamanda en büyük F değerine sahip değişkenin de seçimi olur.

3. Geriye Doğru Eleme (Backward Elimination) Yöntemi: İleriye doğru seçim yönteminin tersine, bu yöntemde önce bütün bağımsız değişkenler modele alınır. Daha sonra belirli kriterlere göre eleme yapılır. SPSS paket programında eleme işlemi için iki ölçüt vardır;

(a) Değişkenin modelde kalabilmesi için sahip olması gereken en küçük F değeri (FOUT).

(b) En büyük F ihtimali (POUT)

En küçük kısmi korelasyon katsayısına sahip olan değişken incelenir. Öngörülen değerlerden daha büyük bir değere sahip olan değişken elenir.

4. Adım Adım Seçme (Stepwise Selection) Yöntemi: İlk bağımsız değişken seçilir. Eğer bu bağımsız değişken ileriye doğru seçim yöntemindeki FIN veya PIN gerekliliklerini yerine getirirse, ikinci değişken seçilir. Aksi halde işlem biter. İkinci değişken olarak en yüksek kısmi korelasyona sahip olan değişken alınır. Değişken seçim işlemi en yüksek kısmi korelasyona sahip değişken alınır. Seçimler yüksek korelasyondan başlanarak en küçüğüne doğru yapılır. Bağımsız değişkenler anlatılan kriterlere uyarsa, regresyon analizine başlanır.

İlk değişken seçildikten sonra adım adım seçme, ileriye doğru seçmeden farklılaşır. İlk değişkenin, geriye doğru eleme yönteminde ki gibi FOUT ya da POUT kriterlerine uyup uymadığı kontrol edilir. Her adımda, modeldeki bağımsız değişkenler eleme açısından kontrol edilir. Başka bir deyişle, adım adım seçim yönteminde, hem ileriye doğru seçim hem de geriye doğru eleme işlemleri yapılır.(9)

3. MATERYAL VE METOD

Çalışmamızda kullanılan veriler, İnönü Üniversitesi Tıp Fakültesi Kardiyoloji Anabilim dalından alınmıştır. Toplam denek sayısı 237 olup, 124'ü Koroner Arter Hastası iken, 113'ü Koroner Arter hastası değildir.

Çalışmamızda, Koroner Arter hastalığını açıklamak için kullanabileceğimiz değişkenler şunlardır:

Kesikli değişkenler:

1. Cinsiyet: Kesikli bir değişken olup, 0: Kadın, 1 : Erkek olarak kodlanmıştır.
2. Hipertansiyon(HT): Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.
3. Diabet Mellitus: Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.
4. Aile Öyküsü: Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.
5. Sigara: Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.
6. Stres: Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.
7. Fiziksel Aktivite: Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.
8. Obezite: Kesikli bir değişken olup, 0: Yok, 1 : Var olarak kodlanmıştır.

Sürekli değişkenler:

1. Yaş : Sürekli bir değişken olup, Yıl olarak ölçülmüştür.
2. Hemoglobin(HB): Sürekli bir değişken olup, gr / dl olarak ölçülmüştür.
3. Beyaz Küre(BK): Sürekli bir değişken olup, ml'de bulunan sayı olarak tespit edilmiştir.
4. Ürik Asit (UA): Sürekli bir değişken olup, gr / dl olarak ölçülmüştür.
5. LDL – Kolesterol : Sürekli bir değişken olup, gr / dl olarak ölçülmüştür.
6. HDL – Kolesterol : Sürekli bir değişken olup, gr / dl olarak ölçülmüştür.
7. Trigliserit(TG): Sürekli bir değişken olup, gr / dl olarak ölçülmüştür.

Yukarıdaki bağımsız (açıklayıcı) değişkenler, Koroner Arter Hastalığına neden olan risk faktörlerini tespit etmek için kullanılabilecek olan değişkenlerdir.

Lojistik regresyon analizinde risk yaklaşımı söz konusu olduğundan sürekli değişkenlerin kendi özgün değerleriyle kullanılması, araştırmacıların amacına göre regresyon katsayılarının yorumlanması işleminde sorun olabilmektedir. Risk hesaplamalarında, riskin hesaplanması için riskin hangi gruba göre yapılacağı tespit edilmelidir. Bu nedenle sürekli değişkenler araştırmacının amacına göre sınıflandırılabilir.(5) Bizde sürekli değişkenleri, İnönü Üniversitesi Tıp Fakültesi Kalp Damar Cerrahisi Anabilim Dalı Araş. Gör. Dr. M. Cengiz Çolak tarafından belirlenen kesim noktalarına göre riskli ve risksiz gruba ayırdık. Şimdi değişkenlere ilişkin kesim noktaları ve kodlanmaları aşağıda verilmiştir.

Tablo 5. Yaş değişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
YAŞ ≤ 60	0
YAŞ > 60	1

Burada riskli grup 60 yaşın üzeri (YAŞ > 60) olarak belirlenmiştir.

Tablo 6. HB değişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
HB ≥ 12	0
HB < 12	1

Burada riskli grup, HB < 12 olarak belirlenmiştir.

Tablo 7. BK değişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
BK < 8000	0
BK ≥ 8000	1

Burada riskli grup, BK ≥ 8000 olarak belirlenmiştir.

Tablo 8. UA deęişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
UA < 7,0	0
UA ≥ 7,0	1

Burada riskli grup UA ≥ 7,0 olarak belirlenmiştir.

Tablo 9. LDL deęişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
LDL ≤ 145	0
LDL > 145	1

Burada riskli grup LDL > 145 olarak belirlenmiştir.

Tablo 10. TG deęişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
TG ≤ 160	0
TG > 160	1

Burada riskli grup TG > 160 olarak belirlenmiştir.

Tablo 11. HDL deęişkenine ilişkin kesim noktaları ve kodlanması

KESİM NOKTASI	KOD
HDL ≥ 30	0
HDL < 30	1

Burada riskli grup HDL < 30 olarak belirlenmiştir.

Verilerin analizinde ilk olarak ileriye doğru deęişken seçimi (Forward Selection) yöntemi kullanılmıştır. İleriye doğru deęişken seçimi (Forward Selection) yönteminde, SPSS paket programı her adımda deęişkenin önemini test edip, belli bir kurala göre deęişkenleri modele dahil etmekte veya çıkarmaktadır.

Daha sonra geriye doğru eleme (Backward Elimination) yöntemi kullanılmıştır. Bu yöntemde, SPSS paket programı tarafından her adımda deęişkenlerin önemlilięi kontrol edip, belli bir kurala göre deęişkenler modelde kalmakta veya çıkarılmaktadır.

Deęişkenler seçildikten sonra, modelin uyum iyilięi testi yapılmıştır. Odds oranları yorumlanıp, modelin ayırımsama gücü ile birlikte seçicilik (specifity) oranı ve duyarlılık (sensitivity) hesaplanmıştır. Son olarak, kullanılan iki yöntemin sonuçları karşılaştırılarak en iyi lojistik model elde edilmeye çalışılmıştır.

Bu çalışmada, SPSS for Windows 8.0 paket programı kullanılmıştır.

4. BULGULAR

Aşağıdaki tablolarda çalışmamıza alınan 237 kişiye ilişkin bazı bilgiler verilmiştir.

Tablo 12. KAH için risk faktörü olabilecek kesikli değişkenlerin KAH olan ve olmayanlara göre Sayı (n) ve Yüzde (%) ile gösterimi.

Değişkenler	KAH(-)		KAH(+)		Toplam	
	n	%	n	%	n	%
Cinsiyet						
Kadın	36	48,6	38	51,4	74	100,0
Erkek	77	47,2	86	52,8	163	100,0
Diabet						
Yok	91	59,1	63	40,9	154	100,0
Var	22	26,5	61	73,5	83	100,0
Diabet						
Yok	92	59,1	63	40,9	155	100,0
Var	21	26,5	61	73,5	82	100,0
Stres						
Yok	54	79,4	14	20,6	68	100,0
Var	59	34,9	110	65,1	169	100,0
F. Aktivite						
Yok	84	41,2	120	58,8	204	100,0
Var	29	87,9	4	12,1	33	100,0
Obezite						
Yok	90	58,8	63	41,2	153	100,0
Var	23	27,4	61	72,6	84	100,0
H.tansiyon						
Yok	90	60,8	58	39,2	148	100,0
Var	23	25,8	66	74,2	89	100,0
Aile öyküsü						
Yok	95	57,6	70	42,4	165	100,0
Var	18	25,0	54	75,0	72	100,0

Tablo 13. KAH için risk faktörü olabilecek sürekli değişkenlerin KAH olan ve olmayanlara göre Sayı (n) ve Yüzde (%) ile gösterimi.

Değişkenler	KAH(-)		KAH(+)		Toplam	
	n	%	n	%	n	%
Yaş						
Yaş ≤ 60	103	60,9	66	39,1	169	100,0
Yaş > 60	10	14,7	58	85,3	68	100,0
Hb						
Hb ≥ 12	101	49,0	105	51,0	206	100,0
Hb < 12	12	38,7	19	61,3	31	100,0
Bk						
Bk < 8000	108	55,0	90	45,0	198	100,0
Bk ≥ 8000	5	12,8	34	87,2	39	100,0
UA						
Ua < 7,0	112	54,0	97	46,0	209	100,0
Ua ≥ 7,0	1	3,6	27	96,4	28	100,0
HDL-Kol.						
Hdl ≥ 30	100	47,8	109	52,2	209	100,0
Hdl < 30	13	46,4	15	53,6	28	100,0
LDL-Kol.						
Ldl ≤ 145	106	58,2	76	41,8	182	100,0
Ldl > 145	7	12,7	48	87,3	55	100,0
Trigliserit						
Tg ≤ 160	105	69,1	47	30,9	152	100,0
Tg > 160	8	9,4	77	90,6	85	100,0

5. TARTIŞMA

Değişken seçim işlemimizde ilk olarak Forward Selection (ileriye doğru değişken seçim) metodunu kullanacağız. Bu metod daha önce anlatıldığı gibi, SPSS paket programı her adımda değişkenin önemini test edip, belli bir kurala göre değişkenleri modele dahil etmekte veya çıkarmaktadır. Şimdi uygulamamıza ilişkin sonuçlar aşağıdaki gibidir.

BİRİNCİ ADIM : İlk olarak TG girecektir. Modelde sadece sabit terim varken, yani başlangıçta -2 Log Likelihood değerimiz, 328,04 idi.

Tablo 14. TG değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIĞI
	-0,80	0,17	20,97	1	0,0000		
	3,068	0,41	55,77	1	0,0000	21,50	9,61-48,1

-2 Log Likelihood = 241,05

Tablo 15. TG değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	105	8	113	92,92
HASTA	47	77	124	62,10
TOPLAM	152	85	237	76,79

İKİNCİ ADIM: Modele YAŞ girecektir.

Tablo 16. YAŞ değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

ŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIĞI
T	-1,48	0,23	39,85	1	0,0000		
	3,26	0,44	54,82	1	0,0000	26,15	11,02-62,06
	2,47	0,43	32,99	1	0,0000	11,83	5,09-27,51

-2 Log Likelihood = 200,65

Tablo 17. YAŞ değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	96	17	113	84,96
HASTA	21	103	124	83,06
TOPLAM	117	120	237	83,97

ÜÇÜNCÜ ADIM: Modele SİGARA girecektir.

Tablo 18. SİGARA değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

İŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIĞI
	2,34	0,46	25,61	1	0,0000	10,46	4,21-25,97
RA	1,78	0,38	21,22	1	0,0000	5,93	2,78-12,65
	3,15	0,47	45,11	1	0,0000	23,53	9,36-59,15
T	-2,32	0,33	46,81	1	0,0000		

-2 Log Likelihood = 177,59

Tablo 19. SİGARA değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	96	17	113	84,96
HASTA	21	103	124	83,06
TOPLAM	117	120	237	83,97

DÖRDÜNCÜ ADIM: Modele H. Tansiyon girecektir.

Tablo 20. H.TANSİYON değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN %95 GÜVEN ARALIĞI
	2,55	0,5173	24,35	1	0,0000	12,84	4,65-36,40
TANSİYON	1,90	0,44	17,94	1	0,0000	6,70	2,77-16,15
RA	2,00	0,42	22,30	1	0,0000	7,40	3,22-16,99
	3,15	0,49	40,70	1	0,0000	23,52	8,91-62,08
T	-3,19	0,45	49,84	1	0,0000		

-2 Log Likelihood = 156,98

Tablo 21. H.TANSİYON değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	99	14	113	87,61
HASTA	20	104	124	83,87
TOPLAM	119	118	237	85,65

BEŞİNCİ ADIM: Modele LDL girecektir.

Tablo 22. LDL değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIĞI
Ş	2,60	0,56	20,98	1	0,0000	13,39	4,41-40,66
ANSİYON	2,12	0,48	19,03	1	0,0000	8,32	3,21-21,56
ARA	2,25	0,47	22,58	1	0,0000	8,52	3,75-24,11
	2,55	0,61	17,48	1	0,0000	12,82	3,87-42,38
	2,99	0,54	30,76	1	0,0000	19,90	6,91-57,26
İT	-3,91	0,55	49,46	1	0,0000		

-2 Log Likelihood = 135,10

Tablo 23. LDL değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	96	17	113	84,96
HASTA	11	113	124	91,13
TOPLAM	107	130	237	88,2

ALTINCI ADIM : Modele AİLE ÖYKÜSÜ girecektir.

Tablo 24. AİLE ÖYKÜSÜ değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

İŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIĞI
	2,58	0,61	17,73	1	0,0000	13,30	3,98-44,32
NSİYON	2,75	0,57	23,25	1	0,0000	15,74	5,13-48,30
ÖYKÜSÜ	2,46	0,63	15,10	1	0,0001	11,73	3,39-40,61
RA	2,55	0,53	23,12	1	0,0000	12,83	4,53-36,30
	3,02	0,70	18,57	1	0,0000	20,63	5,20-81,71
	2,81	0,57	23,85	1	0,0000	16,62	5,38-51,34
T	-5,02	0,73	47,47	1	0,0000		

-2 Log Likelihood = 116,457

Tablo 25. AİLE ÖYKÜSÜ değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	94	19	113	83,18
HASTA	7	117	124	94,35
TOPLAM	101	136	237	89,03

DEĞİŞKENLER	β	S. HA A	WALD	S.	P	ORANI	95% GÜVEN ARALIĞI
	2,76	0,67	17,03	1	0,0000	15,87	4,27-59,02
ANSİYON	2,95	0,64	21,51	1	0,0000	19,16	5,50-66,73
ÖYKÜSÜ	3,03	0,73	17,17	1	0,0000	20,70	4,93-86,73
RA	2,81	0,58	22,98	1	0,0000	16,72	5,28-52,90
	3,00	0,76	15,53	1	0,0001	20,10	4,52-89,36
	2,94	0,63	21,42	1	0,0000	18,98	5,45-66,04
İTE	2,04	0,60	11,91	1	0,0006	7,73	2,42-24,70
T	-6,17	0,93	43,74	1	0,0000		

-2 Log Likelihood = 102,414

Tablo 27. OBEZİTE değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	108	5	113	95,58
HASTA	17	107	124	86,29
TOPLAM	125	112	237	90,72

SEKİZİNCİ ADIM: Modele DİABET girecektir.

Tablo 28. DİABET değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % GÜVEN ARALIĞI
AŞ	3,10	0,74	17,2	1	0,0000	21,99	5,10-94,81
ABET	2,23	0,76	8,42	1	0,0037	9,30	2,06-41,85
TANSİYON	3,43	0,73	21,71	1	0,0000	30,87	7,30-130,68
LE ÖYKÜSÜ	3,26	0,84	14,86	1	0,0001	26,18	4,97-137,75
GARA	3,03	0,65	21,11	1	0,0000	20,73	5,69-75,57
OL	2,83	0,82	11,76	1	0,0006	17,04	3,37-86,15
	2,77	0,64	18,50	1	0,0000	16,06	4,53-56,94
BEZİTE	2,28	0,65	12,17	1	0,0005	9,85	2,72-35,60
BİT	-7,08	1,13	39,23	1	0,0000		

-2 Log Likelihood = 92,41

Tablo 29. DİABET değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	106	7	113	93,81
HASTA	12	112	124	90,32
TOPLAM	118	119	237	91,98

DOKUZUNCU ADIM: Modele STRES girecektir.

Tablo 30. STRES deęişkeni modele dahil edildięinde deęişkenlere iliřkin sonuçlar

ŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIęI
	2,95	0,81	13,29	1	0.0003	19,28	3,93-94,65
ET	2,30	0,82	7,81	1	0.0052	10,01	1,99-50,35
NSİYON	3,49	0,79	19,36	1	0.0000	32,90	6,94-155,95
ÖYKÜSÜ	3,06	0,85	12,87	1	0.0003	21,50	4,02-114,95
RA	2,94	0,68	18,63	1	0.0000	19,00	4,99-72,34
	2,72	0,86	9,97	1	0.0016	15,21	2,80-82,33
	2,76	0,66	17,44	1	0.0000	15,88	4,34-58,16
S	1,81	0,70	6,64	1	0.0099	6,13	1,54-34,34
İTE	2,61	0,72	12,87	1	0.0003	13,65	3,27-56,95
	-8,43	1,39	36,43	1	0.0000		

-2 Log Likelihood = 84,83

Tablo 31. STRES deęişkeni modele dahil edildięinde ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEęERLER

GÖZLENEN DEęERLER	HASTA DEęİL	HASTA	TOPLAM	DOęRU SINIFLANDIRMA YÜZDESİ
HASTA DEęİL	107	6	113	94,7
HASTA	13	111	124	89,5
TOPLAM	120	117	237	91,98

ONUNCU ADIM: Modele BEYAZ KÜRE girecektir. Bu adımda işlem durdurulur. Sonuç olarak modeldeki değişkenler Tablo 32'de verilmiştir.

Tablo 32. BEYAZ KÜRE değişkeni modele dahil edildiğinde değişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN %95 GÜVEN ARALIĞI
BAŞ	2,79	0,83	11,15	1	0,0008	16,30	3,17-83,86
ABET	2,99	0,94	10,18	1	0,0014	20,06	3,18-126,58
FANSİYON	4,07	0,91	19,85	1	0,0000	58,86	9,80-353,51
LE ÖYKÜSÜ	3,18	0,89	12,64	1	0,0004	24,28	4,18-140,94
BARA	3,16	0,74	18,15	1	0,0000	23,60	5,51-101,07
KÜRE	2,84	1,11	6,47	1	0,0109	17,12	1,92-152,70
L	2,62	0,87	8,98	1	0,0027	13,81	2,48-76,86
	2,27	0,68	11,08	1	0,0009	9,75	2,55-37,30
RES	2,20	0,79	7,76	1	0,0053	9,10	1,92-43,08
EZİTE	2,57	0,80	10,24	1	0,0014	13,14	2,71-63,60
BİT	-9,45	1,65	32,83	1	0,0000		

-2 Log Likelihood = 76,613

Tablo 33. BEYAZ KÜRE değişkeni modele dahil edildiğinde ortaya çıkan sınıflandırma tablosu

GÖZLENEN DEĞERLER	TAHMİN EDİLEN DEĞERLER			DOĞRU SINIFLANDIRMA YÜZDESİ
	HASTA DEĞİL	HASTA	TOPLAM	
HASTA DEĞİL	107	6	113	94,7
HASTA	9	115	124	92,7
TOPLAM	116	121	237	93,7

İleriye doğru deęişken seçim yöntemi sonucunda modelde bulunan bağımsız deęişkenler şunlardır:

- (1) YAŞ
- (2) DİABET
- (3) H.TANSİYON
- (4) A.ÖYKÜSÜ
- (5) SİGARA
- (6) LDL-KOLESTEROL
- (7) TRIGLİSERİT
- (8) STRES
- (9) OBEZİTE
- (10) B. KÜRE

Buna karşın,

- (1) CİNSİYET
- (2) HEMOGLOBİN
- (3) ÜRİK ASİT
- (4) HDL-KOLESTEROL
- (5) FİZİK AKTİVİTE

bağımsız deęişkenleri ileriye doğru deęişken seçim yöntemi sonucunda modele alınmamışlardır.

Uyum iyiliği testi ise; SPSS paket programı tarafından hesaplanan Hosmer – Lemeshow'un uyum iyiliği testi ile yapılmış olup, sonuçlar şöyledir:

Tablo 34. Hosmer – Lemeshow uyum iyiliği testi sonuçları

UYUM İYİLİĞİ TESTİ	Kİ-KARE	SERBESTLİK DERECEŚİ	P
HOSMER-LEMESHOW	2,03	8	0,98

Görüldüğü gibi $P=0,98$ oldukça büyük olup, elde edilen modelin oldukça uyumlu olduğu sonucu elde edilmektedir.

Lojistik regresyon analizinin ayırimsama gücü yukarıdaki tablo 33'de verilmiştir. Modelin toplam ayırimsama gücünün % 93,7 olduğu görülmektedir. Koroner arter hastalığına sahip olmayanların doğru olarak sınıflandırılma (duyarlılık; sensitivity) oranı % 94,7 olarak tespit edildi.

Ayrıca Koroner arter hastalığına sahip olanların doğru olarak sınıflandırılma (seçicilik; specificity) oranı % 92,7 olarak tespit edildi.

Şimdi elde edilen odds oranlarını (Tablo 32'ye göre) inceleyelim.

- 60 yaşın üzerindeki ($YAŞ>60$) kişilerde koroner arter hastalığına sahip olma riskinin, 60 yaş ve altındaki ($YAŞ\leq 60$) kişilere göre 16,3 kat daha fazla olduğu tespit edilmiştir.
- Diabeti olan kişilerde koroner arter hastalığına sahip olma riskinin, Diabeti olmayanlara göre 20,06 kat daha fazla olduğu tespit edilmiştir.
- H. Tansiyonu olan kişilerde koroner arter hastalığına sahip olma riskinin, H. Tansiyonu olmayanlara göre 58,86 kat daha fazla olduğu tespit edilmiştir.

- Aile öyküsü olan kişilerde koroner arter hastalığına sahip olma riskinin, Aile öyküsü olmayanlara göre 24,3 kat daha fazla olduğu tespit edilmiştir.
- Sigara içen kişilerde koroner arter hastalığına sahip olma riskinin, Sigara içmeyenlere göre 23,6 kat daha fazla olduğu tespit edilmiştir.
- LDL-Kolesterolü 145'den fazla ($LDL > 145$) olan kişilerde koroner arter hastalığına sahip olma riskinin, LDL-Kolesterolü 145 ve daha az ($LDL \leq 145$) olanlara göre 13,8 kat daha fazla olduğu tespit edilmiştir.
- Trigliseriti 160'dan fazla ($TG > 160$) olan kişilerde koroner arter hastalığına sahip olma riskinin, Trigliseriti 160 ve daha az ($TG \leq 160$) olanlara göre 9,75 kat daha fazla olduğu tespit edilmiştir.
- Stresi olan kişilerde koroner arter hastalığına sahip olma riskinin, stresi olmayanlara göre 9,1 kat daha fazla olduğu tespit edilmiştir.
- Obezitesi olan kişilerde koroner arter hastalığına sahip olma riskinin, Obezitesi olmayanlara göre 13,14 kat daha fazla olduğu tespit edilmiştir.
- Beyaz küresi 8800 ve daha fazla ($BK \geq 8800$) olan kişilerde koroner arter hastalığına sahip olma riskinin, Beyaz küresi 8800'den küçük ($BK < 8800$) olanlara göre 17,12 kat daha fazla olduğu tespit edilmiştir.

Forward Selection (ileriye doğru) deęişken seçme yönteminden sonra, şimdi Backward Elimination (geriye doğru) deęişken eleme yöntemini kullanalım. Birinci adımda bütün deęişkenler modele alınır, sonraki adımlarda SPSS paket programı her adımda deęişkenlerin önemlilięi kontrol edip, belli bir kurala göre deęişkenler modelde kalmakta veya çıkarılmaktadır. Şimdi uygulamamıza ilişkin sonuçlar aşağıdaki gibidir:

BİRİNCİ ADIM: Modele bütün deęişkenler alınır. Deęişkenler şunlardır:

CİNS, YAŞ, DIABET, HT, AÖ, SİGARA, HB, BK, UA, HDL, LDL, TG, STRES, F.AKTİVİTE, OBEZİTE. Başlangıçta $-2 \text{ Log Likelihood}$ deęerimiz, 328,041'dir.

Tablo 35. Bütün deęişkenler modelde iken deęişkenlere ilişkin sonuçlar

DEęİŐKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P
CİNS	-0,8658	0,78	1,22	1	0,2690
YAŞ	2,71	0,97	7,74	1	0,0554
DIABET	3,09	1,03	9,01	1	0,0027
H.TANSİYON	4,43	1,06	17,36	1	0,0000
AİLE ÖYKÜSÜ	4,01	1,08	13,64	1	0,0002
SİGARA	3,68	0,93	15,66	1	0,0001
HB	2,16	1,24	3,03	1	0,0813
BK	1,71	1,27	1,80	1	0,1793
UA	3,36	1,92	3,07	1	0,0795
HDL	0,96	1,09	0,77	1	0,3781
LDL	3,11	1,07	8,37	1	0,0038
TG	2,33	0,82	7,99	1	0,0047
STRES	3,05	0,99	9,47	1	0,0021
F.AKTİVİTE	0,53	1,07	0,24	1	0,6172
OBEZİTE	2,88	0,85	9,11	1	0,0025
SABİT	-10,89	2,23	23,79	1	0,0000

$-2 \text{ Log Likelihood} = 67,778$

Tablo 36. Bütün deęişkenler modelde iken ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	107	6	113	94,7
HASTA	9	115	124	92,7
TOPLAM	116	121	237	93,7

İKİNCİ ADIM: Modelden F. AKTİVİTE çıkarılır. Bu adımda, -2 Log Likelihood = 68,02.

Tablo 37. F. aktivite deęişkeni modelden çıkarıldığında modeldeki deęişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P
CİNS	-0,89	0,78	1,30	1	0,2536
YAŞ	2,61	0,95	7,58	1	0,0059
DİABET	3,03	1,02	8,83	1	0,0030
H.TANSİYON	4,29	1,01	18,04	1	0,0000
AİLE ÖYKÜSÜ	3,91	1,05	13,75	1	0,0002
SİGARA	3,66	0,92	15,62	1	0,0001
HB	2,16	1,26	2,93	1	0,0869
BK	1,62	1,26	1,64	1	0,1994
UA	3,52	1,99	3,12	1	0,0772
HDL	1,05	1,09	0,92	1	0,3362
LDL	2,97	1,01	8,56	1	0,0034
TG	2,23	0,79	7,94	1	0,0048
STRES	3,02	1,00	9,08	1	0,0026
OBEZİTE	2,84	0,95	8,94	1	0,0028
SABİT	-10,59	2,13	24,70	1	0,0000

Tablo 38. F. aktivite deęişkeni modelden çıkarıldığında ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	107	6	113	94,7
HASTA	9	115	124	92,7
TOPLAM	116	121	237	93,7

ÜÇÜNCÜ ADIM: Modelden HDL çıkarılır. Bu adımda, -2 Log Likelihood = 68,95'dir.

Tablo 39. HDL deęişkeni modelden çıkarıldığında modeldeki deęişkenlere ilişkin sonuçlar

DEĞİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P
CİNS	-0,79	0,76	1,08	1	0,2985
YAŞ	2,54	0,90	7,92	1	0,0049
DİABET	2,85	0,96	8,66	1	0,0032
H.TANSİYON	4,12	0,97	17,98	1	0,0000
AİLE ÖYKÜSÜ	3,79	1,00	14,20	1	0,0002
SİGARA	3,39	0,82	16,79	1	0,0000
HB	2,08	1,28	2,63	1	0,1047
BK	1,64	1,28	1,65	1	0,1985
UA	3,75	2,25	2,76	1	0,0965
LDL	2,95	1,02	8,27	1	0,0040
TG	2,27	0,79	8,13	1	0,0043
STRES	2,90	0,99	8,56	1	0,0034
OBEZİTE	2,72	0,89	9,20	1	0,0024
SABİT	-10,12	1,99	25,74	1	0,0000

Tablo 40. HDL deęişkeni modelden çıkarıldığında ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEęERLER

GÖZLENEN DEęERLER	HASTA DEęİL	HASTA	TOPLAM	DOęRU SINIFLANDIRMA YÜZDESİ
HASTA DEęİL	107	6	13	94,7
HASTA	9	115	124	92,7
TOPLAM	116	121	237	93,7

DÖRDÜNCÜ ADIM: Modelden CİNSİYET çıkarılır. Bu adımda, -2 Log Likelihood, 70,05.

Tablo 41. CİNSİYET deęişkeni modelden çıkarıldığında modeldeki deęişkenlere ilişkin sonuçlar

DEęİŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P
YAŞ	2,49	0,89	7,77	1	0,0053
DİABET	3,01	0,95	10,05	1	0,0015
H.TANSİYON	4,09	0,95	18,52	1	0,0000
AİLE ÖYKÜSÜ	3,61	0,97	13,79	1	0,0002
SİGARA	3,13	0,77	16,52	1	0,0000
HB	2,23	1,23	3,30	1	0,0692
BK	1,80	1,28	1,96	1	0,1612
UA	3,42	2,10	2,64	1	0,1038
LDL	2,88	1,00	8,28	1	0,0040
TG	2,35	0,77	9,10	1	0,0025
STRES	2,88	0,97	8,72	1	0,0031
OBEZİTE	2,81	0,89	9,81	1	0,0017
SABİT	-10,48	1,96	28,49	1	0,0000

Tablo 42. CİNSİYET değişkeni modelden çıkarıldığında ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER				
GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	108	5	113	95,58
HASTA	8	116	124	93,55
TOPLAM	116	121	237	94,51

BEŞİNCİ ADIM: Modelden BK çıkarılır. İşlem bu adımda durdurulur. Daha hiç bir değişken modelden çıkarılmaz. Sonuç olarak modeldeki değişkenler Tablo 43'te verilmiştir. Bu adımda, -2 Log Likelihood = 72,266'dır.

Tablo 43. BK değişkeni modelden çıkarıldığında modeldeki değişkenlere ilişkin sonuçlar

ŞKENLER	$\hat{\beta}$	S. HATA	WALD	S.DERECESİ	P	ODDS ORANI	OR İÇİN % 95 GÜVEN ARALIĞI
	2,63	0,89	8,70	1	0,0032	13,91	2,41-80,10
ET	2,74	0,92	8,87	1	0,0029	15,56	2,55-94,67
NSİYON	3,90	0,92	17,98	1	0,0000	49,80	8,18-303,18
ÖYKÜSÜ	3,70	0,96	14,61	1	0,0001	40,54	6,07-270,18
RA	3,02	0,74	16,57	1	0,0000	20,61	4,80-38,47
	2,87	1,13	6,44	1	0,0111	17,78	1,92-163,99
	3,70	1,76	4,37	1	0,0366	40,42	1,25-1297,4
	2,84	0,99	8,14	1	0,0043	17,25	2,43-122,01
	2,77	0,73	14,20	1	0,0002	16,04	3,78-67,94
S	2,94	0,97	9,18	1	0,0024	18,92	2,82-126,71
İTE	2,91	0,86	11,34	1	0,0008	18,38	3,37-100,10
Γ	-10,42	1,93	28,99	1	0,0000		

Tablo 44. BK deęişkeni modelden çıkarıldığında ortaya çıkan sınıflandırma tablosu

TAHMİN EDİLEN DEĞERLER

GÖZLENEN DEĞERLER	HASTA DEĞİL	HASTA	TOPLAM	DOĞRU SINIFLANDIRMA YÜZDESİ
HASTA DEĞİL	109	4	113	96,46
HASTA	9	115	124	92,74
TOPLAM	118	119	237	94,51

Şimdi her adımda yapılan işlemleri özetleyen Tablo 45'yi verelim.

Tablo 45. Geriye doğru deęişken seçimi metodunun adımlarına ilişkin sonuçlar

ADIM	MODELDE ÇIKARILAN DEĞİŞKEN	DOĞRU SINIFLANDIRMA YÜZDESİ
1	YOK	93,7
2	F.AKTİVİTE	93,7
3	HDL	93,7
4	CİNS	94,5
5	BK	94,5

Uyum iyiliği testi ise; SPSS paket programı tarafından hesaplanan Hosmer – Lemeshow'un uyum iyiliği testi ile yapılmış olup, sonuçlar şöyledir:

Tablo 46. Hosmer – Lemeshow uyum iyiliği testi sonuçları

UYUM İYİLİĞİ TESTİ	Kİ-KARE	SERBESTLİK DERECEŚİ	P
HOSMER-LEMESHOW	5,48	8	0,70

Görüldüğü gibi $P=0,70$ oldukça büyük olup, elde edilen modelin oldukça uyumlu olduğu sonucu elde edilmektedir.

Lojistik regresyon analizinin ayırimsama gücü yukarıdaki Tablo 44'de verilmiştir. Modelin toplam ayırimsama gücünün % 94,51 olduğu görülmektedir. Koroner arter hastalığına sahip olmayanların doğru olarak sınıflandırılma (duyarlılık; sensitivity) oranı % 96,46 olarak tespit edildi.

Ayrıca Koroner arter hastalığına sahip olanların doğru olarak sınıflandırılma (seçicilik; specificity) oranı % 92,74 olarak tespit edildi.

Şimdi elde edilen odds oranlarını (Tablo 43'e göre) inceleyelim.

- 60 yaşın üzerindeki ($YAŞ > 60$) kişilerde koroner arter hastalığına sahip olma riskinin, 60 yaş ve altındaki ($YAŞ \leq 60$) kişilere göre 13,9 kat daha fazla olduğu tespit edilmiştir.
- Diabeti olan kişilerde koroner arter hastalığına sahip olma riskinin, Diabeti olmayanlara göre 15,5 kat daha fazla olduğu tespit edilmiştir.
- H. Tansiyonu olan kişilerde koroner arter hastalığına sahip olma riskinin, H. Tansiyonu olmayanlara göre 49,8 kat daha fazla olduğu tespit edilmiştir.

- Aile öyküsü olan kişilerde koroner arter hastalığına sahip olma riskinin, Aile öyküsü olmayanlara göre 40,5 kat daha fazla olduğu tespit edilmiştir.
- HB değeri 12'den küçük ($HB < 12$) olan kişilerde koroner arter hastalığına sahip olma riskinin, HB değeri 12'den büyük ve eşit ($HB \geq 12$) olanlara göre 17,7 kat daha fazla olduğu tespit edilmiştir.
- UA değeri 7,0'dan büyük ve eşit ($UA \geq 7,0$) olan kişilerde koroner arter hastalığına sahip olma riskinin, UA değeri 7,0'dan küçük ($UA < 7,0$) olanlara göre 40,4 kat daha fazla olduğu tespit edilmiştir.
- Sigara içen kişilerde koroner arter hastalığına sahip olma riskinin, Sigara içmeyenlere göre 20,6 kat daha fazla olduğu tespit edilmiştir.
- LDL-Kolesterolü 145'den fazla ($LDL > 145$) olan kişilerde koroner arter hastalığına sahip olma riskinin, LDL-Kolesterolü 145 ve daha az ($LDL \leq 145$) olanlara göre 17,2 kat daha fazla olduğu tespit edilmiştir.
- Trigliseriti 160'dan fazla ($TG > 160$) olan kişilerde koroner arter hastalığına sahip olma riskinin, Trigliseriti 160 ve daha az ($TG \leq 160$) olanlara göre 16,04 kat daha fazla olduğu tespit edilmiştir.
- Stresi olan kişilerde koroner arter hastalığına sahip olma riskinin, stresi olmayanlara göre 18,9 kat daha fazla olduğu tespit edilmiştir.
- Obesitesi olan kişilerde koroner arter hastalığına sahip olma riskinin, Obesite'si olmayanlara göre 18,38 kat daha fazla olduğu tespit edilmiştir.

Şimdi ileriye doğru seçim yöntemi ile geriye doğru değişken seçimi yöntemi sonuçlarını karşılaştıralım.

Tablo 47. İleriye doğru değişken seçim yöntemi ile geriye doğru değişken eleme yönteminin karşılaştırılması

İLERİYE DOĞRU DEĞİŞKEN SEÇİM YÖNTEMİ	GERİYE DOĞRU DEĞİŞKEN ELEME YÖNTEMİ
YAŞ	YAŞ
DİABET	DİABET
H.TANSİYON	H.TANSİYON
AİLE ÖYKÜSÜ	AİLE ÖYKÜSÜ
SİGARA	SİGARA
B.KÜRE	HB
LDL	UA
TG	LDL
STRES	TG
OBEZİTE	STRES
	OBEZİTE

Tablo 48. İleriye doğru değişken seçim yöntemi ile geriye doğru değişken eleme yönteminin -2 Log Likelihood'ları

	İLERİYE DOĞRU DEĞİŞKEN SEÇİM YÖNTEMİ	GERİYE DOĞRU DEĞİŞKEN ELEME YÖNTEMİ
-2 LOG LIKELIHOOD	76,613	72,266

Tablo 48'den G istatistiği, 4,37 olarak bulunur. Burada serbestlik derecesi, iki modelde bulunan değişken sayılarının farkı olan 1'e eşittir. $\alpha = 0,05$ ve serbestlik derecesi 1 olan ki-kare tablo değeri, 3,84'dür. Bu durumda $4,37 > 3,84$ olduğundan en iyi lojistik model, Geriye Doğru Değişken Eleme yöntemi ile bulunan model olacaktır.

6. SONUÇLAR

Çalışmamızda, bağımlı değişken olarak Koroner Arter Hastalığı alınmıştır. İleriye doğru değişken seçim yöntemi ve geriye doğru değişken eleme yöntemi sonucunda lojistik modeller bulunmuştur. Elde edilen bu iki modelden G istatistiği hesaplanmış ve test sonucunda geriye doğru değişken eleme yöntemi ile bulunan model en iyi model olarak belirlenmiştir. Elde edilen bu modelden tahmin edilen olasılıklar yardımıyla, Koroner Arter Hastası olan bir kişi % 92,74 ihtimalle doğru olarak sınıflandırılırken, Koroner Arter Hastası olmayan bir kişi % 96,46 ihtimalle doğru olarak sınıflanmıştır. Ayrıca bireylerin ait oldukları gruba ise, % 94,51 gibi yüksek bir yüzde ile atandığı tespit edilmiştir.

Koroner Arter Hastalığı için risk faktörü olabileceği düşünülen 15 değişkenden 11 değişken final modelde yer almıştır. Değişkenlere ilişkin Odds oranları yorumlanmıştır. Son olarak final model için uyum iyiliği testi yapılmış ve modelin iyi bir uyuma sahip olduğu sonucuna varılmıştır.

Final modelimizde Koroner Arter Hastalığı için risk faktörü olarak bulunan değişkenler, diğer yapılan çalışmaların sonuçlarıyla benzer olduğu görülmüştür.(14)

7. ÖZET

Lojistik regresyon analizi ve sađlık bilimlerinde bir uygulama

Çalışmamızda, lojistik regresyon analizi, tahmin yöntemleri, odds oranı, uyum iyiliđi ve deđişken seçim yöntemleri açıklanmıştır.

İleriye dođru deđişken seçim yöntemi ve geriye dođru deđişken eleme yöntemi, kesikli ve sürekli deđişkenlerden oluşan Kardiyolojik verilere uygulanmış ve Koroner arter hastalığını tespit eden en iyi lojistik model bulunmuştur.

Uygulama sonucunda elde edilen sonuçlar yorumlanmıştır.

Anahtar Kelimeler: Logistik regresyon analizi, tahmin yöntemleri, odds oranı, uyum iyiliđi ve deđişken seçim yöntemleri.

8. SUMMARY

Logistic regression analysis and its application in health sciences.

In our study, logistic regression analysis, methods of estimation, Odds Ratio, Goodness of fit and methods of variable selection are explained.

Forward variable selection method and backward elimination method are applied to cardiologic data containing continuous and discrete variables and the best logistic model which determines coronary artery disease have been found.

The results of this application are evaluated.

Key Words: Logistic regression analysis, methods of estimation, Odds Ratio, Goodness of fit and methods of variable selection.

**TC YÖKSEKÖĞRETİM KURULU
FARSAHATASYON MERKEZİ**

9. KAYNAKLAR

- (1) HOSMER, D. W., LEMESHOW, S.: *Aplied Logistic Regression*, John Wiley & Sons, 1989.
- (2) KLEİNBAUM, DAVID G.: *Logistic Regression: A self-Learning Text*, 1992.
- (3) ÖZDAMAR, K.: *Paket Programlar ile İstatistiksel Veri Analizi – 1*, 1999.
- (4) BAŞARIR, G.: *Çok Değişkenli Verilerde Ayırsama Sorunu ve Lojistik Regresyon Analizi*, Hacettepe Üniversitesi, Doktora Tezi, 1990.
- (5) EDİZ, B.: *Lojistik Regresyon: Ayırma Analizi. Doktora Tezi*, 1997.
- (6) ÜNAL, A.: *Lojistik Regresyon Analizi ve Uygulaması*, Gazi Üniversitesi, Yüksek Lisans Tezi, 1997.
- (7) SANFORT, WEISBERG.: *Aplied Linear Regression*, University of Minnesota, 1985.
- (8) ELHAN, ATILLA H.: *Lojistik regresyon analizinin incelenmesi ve tıpta bir uygulaması*, Ankara Üniversitesi, Yüksek Lisans Tezi, 1997.
- (9) TAŞDELEN, A.: *Lojistik regresyon modelinin yaşam verilerine uygulanması*, Ege Üniversitesi, 1997.
- (10) TATLIDİL, H.: *Uygulamalı Çok değişkenli istatistiksel Analiz*, 1996.
- (11) TAŞKIN, V.: *Lojistik regresyon yöntemi ile Angina pectoris tanısına çeşitli ölçütlerin katkısının incelenmesi*, İstanbul üniversitesi, 1997.
- (12) ATA KURT, Y.: *Lojistik regresyon analizi ve Tıp alanında kullanımına ilişkin bir uygulama*, Ankara Üniversitesi, 1999.

- (13) KOÇAK, İ.: Temel bileşenler analizi ve uygulaması, İnönü üniversitesi, Yüksek lisans tezi, 1997.
- (14) ÇEVİRİ EDITÖRLERİ: KEKLİKOĞLU, M., TUZCU, M.: The Merck Manual, Cilt:1-2, İstanbul 1995.
- (15) FREUND, J.R., WILSON, W.J.: Regression Analysis Statistical Modeling of a Response variable, 1998.
- (16) CHRISTENSEN, R.: Log-Linear Models, University of NewMexico, USA, 1990.
- (17) MCCULLAGH, P., NELDER, J.A., CHAPMAN AND HALL: Generalized Linear Models.
- (18) SARAÇBAŞI, O.: Lojistik Regresyon, Lisans üstü eğitim için hazırlanmış ders notu, Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Bilim Dalı, 1994.