



**T.C.**  
**SÜLEYMAN DEMİREL ÜNİVERSİTESİ**  
**SAĞLIK BİLİMLERİ ENSTİTÜSÜ**  
**BİYOİSTATİSTİK VE TIBBİ BİLİŞİM ANABİLİM DALI**

**LOJİSTİK REGRESYON ANALİZİ VE SAĞLIK ALANINDA BİR**  
**UYGULAMA**

**HAZIRLAYAN**  
**Büşra ÇATAL**

**YÜKSEK LİSANS TEZİ**

**DANIŞMAN**  
**Dr. Öğr. Üyesi Osman GÜRDAL**

**ISPARTA-2021**

## KABUL ve ONAY SAYFASI

Sağlık Bilimleri Enstitü Müdürlüğüne;

Süleyman Demirel Üniversitesi Sağlık Bilimleri Enstitüsü **Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı Yüksek Lisans Programı** Çerçevesinde yürütölmüş olan bu çalışma, aşğıdaki jüri tarafından **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Tez Savunma Tarihi: 28/10/2021

Tez Danışmanı : Dr. Öğr. Üyesi Osman GÜRDAL  
Süleyman Demirel Üniversitesi,  
Tıp Faköltesi Biyoistatistik ve Tıbbi Bilişim A.D

Üye : Prof. Dr. Hikmet ORHAN

Üye : Prof. Dr. Soner ÇANKAYA

ONAY: Bu yüksek lisans tezi, Enstitü Yönetim Kurulu'na belirlenen yukarıdaki jüri üyeleri tarafından uygun görölmüş ve kabul edilmiştir.

Prof. Dr. Ebru ÇUBUK DEMİRALAY  
Enstitü Müdürü

## **BEYAN**

*“Lojistik Regresyon Analizi ve Saęlık Alanında Bir Uygulama”* adlı Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Saęlık Bilimleri Enstitüsü Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

**Tezi Hazırlayan**

Büşra ÇATAL

**Danışman**

Dr. Öğr. Üyesi Osman GÜRDAL

## TEŞEKKÜR

Yüksek lisans öğrenimimin başından sonuna kadar bilgi birikimlerini benimle paylaşan, akademik olarak gelişmemde yanımda olan danışman hocam Dr. Öğr. Üyesi Osman GÜRDAL'a, değerli katkılarından dolayı bölüm hocalarımdan Prof. Dr. Hikmet ORHAN'a ve Dr. Öğr. Üyesi Adnan KARAİBRAHİMOĞLU'na teşekkür ve saygılarımı sunarım.

Hayatım boyunca bana olan sevgisini, sabrını, şefkatini bir an olsun esirgemeyen başarılı olmam için bütün desteğiyle yanımda olan annem Hatice BİLTEKİN'e, yüksek lisans sürecinde çalışmalarımı hep sabırla karşılayan eşim Ömer ÇATAL, çocuklarım İbrahim Emir ve Ela ÇATAL'a sonsuz sevgi ve saygılarımı sunarım.

**Büşra ÇATAL**  
**Isparta, 2021**

## ÖZET

### Lojistik Regresyon Analizi ve Sağlık Alanında Bir Uygulama

**Amaç:** Bu çalışmada bağımlı değişkenin sürekli veya kategorik olduğu durumlarda kullanılabilecek lojistik regresyon analizi incelenmiştir. Kullanıma açık verilerden yararlanılarak lojistik regresyon analizinin iki ve çok kategorili modellerinin teorisi ve analizlerinin ayrıntılı bir şekilde anlatılması amaçlanmıştır.

**Yöntem:** Retrospektif olarak yapılan bu araştırmada iki kategorili lojistik regresyon analizi için kalp krizi riski veri seti ve çok kategorili lojistik regresyon analizi için sağlık sigortası harcaması veri seti kullanılmıştır. Verilerin normal dağılım gösterip varyanslarının homojen olması durumunda bağımsız grup Student t-testi ve üç veya daha fazla çoklu grupların karşılaştırılmalarında ise tek yönlü varyans Analizi (Anova) yapılmıştır. Verilerin normal dağılım göstermediği veya kategorik verilerin analizleri için parametrik olmayan Mann-Whitney U testi ve üç veya daha fazla gruplara göre karşılaştırmalar için Kruskal-Wallis H testi uygulanmıştır.

**Bulgular:** Hastaların yaşı, dinlenme anı kan basıncı, maksimum nabız sayısı, ST segmenti çökme değeri ve görüntülenen damar sayısı, düşük ve yüksek MI risk sınıfları arasında anlamlı farklılık göstermiştir. Dinlenme anı EKG bulgularında ST-T dalga anormalliği yüksek risk grubunda bulunan hastalarda daha yüksek oranda gözlenmiştir. Egzersiz anında anjina görülme oranı düşük risk grubundaki hastalarda daha yüksek oranda izlenmiştir. Çok kategorili lojistik regresyon analizinde müşterilerin VKİ değerleri ( $p<0,001$ ) ve ödediği sağlık sigorta miktarı ( $p=0,010$ ) Güneydoğu Bölgesi'nde yaşayanlar için anlamlı düzeyde yüksektir. Müşterilerin ödediği sağlık sigorta miktarı da Güneydoğu Bölgesi müşterilerinde anlamlı düzeyde yüksek bulunmuştur ( $p=0,010$ ).

**Sonuç:** İki kategorili lojistik regresyon analizinde kullanılan kalp krizi verisinde hastaların yaşı, dinlenme anı kan basıncı, maksimum nabız sayısı, ST segmenti çökme değeri ve görüntülenen damar sayısı, düşük ve yüksek MI risk sınıfları arasında anlamlı farklılık göstermiştir ( $p<0,001$ ). İkinci analizde ise, müşterilerin VKİ değerlerinin yüksek olması ve sigara kullanımı sigorta ücretinin önemli bir şekilde artışı modelde yer almaktadır.

**Anahtar Kelimeler:** Lojistik regresyon analizi, iki kategorili, çok kategorili.

# ABSTRACT

## Logistic Regression Analysis and an Application in Healthcare

**Objective:** The increase in the risk of heart attack and the rising health insurance costs in different regions of the States were examined due to different parameters implementing logistic regression. It was aimed to explain the theory and analyses of the dichotomous and multinominal models of the regression in details via publicly available data.

**Method:** In this retrospective study, the heart attack risk dataset for binary logistic regression and the health insurance expenditure dataset for multi-category logistic regression analysis were used. If the data were normally distributed and the variances were homogeneous, the independent groups Student t-test was executed, and one-way analysis of variance (ANOVA) was implemented for the comparison of three or more multiple groups. The non-parametric of Mann-Whitney U test was employed for the analysis of the dataset where there was non-normal distribution, or it was a categorical dataset. When there were three or more groups' comparisons of non-normal or qualitative data, the Kruskal-Wallis H test was carried out for the analysis.

**Results:** The patients' age, resting blood pressure, maximum heart rate, ST segment depression value, and the number of vessels displayed significantly between low and high MI risk classes. In the resting ECG findings, ST-T wave abnormality was observed at a higher rate in patients for the high-risk group. The incidence of angina during exercise was higher in patients for the low-risk group. In the multinominal logistic regression analysis, the BMI values of the customers were significantly higher for those living in the Southeast region ( $p < 0.001$ ). Thus, the amount of health insurance paid by the customers was also observed to be notably higher in the Southeast region customers ( $p = 0.010$ ).

**Conclusion:** In this work, it is presented that the patients' age, resting blood pressure, maximum heart rate, ST segment depression value and the number of displayed vessels show significant differences between low and high MI risk classes ( $p < 0.001$ ). In the second example, the customers who have higher BMI and smoking status that contribute to the insurance cost substantially in the model.

**Keywords:** Logistic regression analysis, binary, multinominal.

# İÇİNDEKİLER

|                                                                           |      |
|---------------------------------------------------------------------------|------|
| BEYAN.....                                                                | i    |
| TEŞEKKÜR .....                                                            | ii   |
| ÖZET .....                                                                | iii  |
| ABSTRACT.....                                                             | iv   |
| İÇİNDEKİLER .....                                                         | v    |
| TABLolar DİZİNİ.....                                                      | vii  |
| ŞEKİLLER DİZİNİ.....                                                      | viii |
| SİMGELER ve KISALTMALAR DİZİNİ.....                                       | ix   |
| 1. GİRİŞ .....                                                            | 1    |
| 2. GENEL BİLGİLER.....                                                    | 3    |
| 2.1. Lojistik Regresyon Analizi .....                                     | 4    |
| 2.1.1. Lojistik Regresyon Analizi Tanımı ve Tarihçesi.....                | 5    |
| 2.1.2. Lojistik Regresyon Analizinin Tercih Nedenleri .....               | 6    |
| 2.1.3. Lojistik Regresyonun Kullanım Alanları .....                       | 7    |
| 2.1.4. Bağımsız Değişkenler.....                                          | 7    |
| 2.1.5. Bağımlı Değişkenlerin Kodlanması .....                             | 8    |
| 2.1.6. Risk Kestirimi.....                                                | 8    |
| 2.1.6.1. Genel Risk Ölçüleri.....                                         | 8    |
| 2.1.6.2. Bağımlı Risk Ölçüleri .....                                      | 9    |
| 2.2. Lojistik Regresyon Çeşitleri .....                                   | 12   |
| 2.2.1. İki Kategorili (Binary) Lojistik Regresyon .....                   | 12   |
| 2.2.1.1. Lojistik Regresyon Modeli.....                                   | 12   |
| 2.2.1.2. Regresyon Katsayılarının Kestirimi.....                          | 14   |
| 2.2.1.2.1. En Çok Olabilirlik Yöntemi.....                                | 14   |
| 2.2.1.2.2. Yeniden Ağırlıklandırılmış En Küçük Kareler Yöntemi.....       | 15   |
| 2.2.1.2.3. Minimum Lojit Ki-Kare Yöntemi .....                            | 15   |
| 2.2.1.3. Katsayıların Önemliliğinin Test Edilmesi.....                    | 16   |
| 2.2.1.3.1. Olabilirlik Oran Testi.....                                    | 16   |
| 2.2.1.3.2. Wald Testi.....                                                | 16   |
| 2.2.1.3.3. Skor Testi.....                                                | 17   |
| 2.2.1.4. Bağımsız Değişkenlerin Lojistik Regresyon Modeline Alınması..... | 17   |

|                                                                        |           |
|------------------------------------------------------------------------|-----------|
| 2.2.1.4.1. Sayısal (Niceliksel) Değişkenler.....                       | 17        |
| 2.2.1.4.2. Niteliksel Değişkenler.....                                 | 18        |
| 2.2.1.5. Çoklu Lojistik Regresyon .....                                | 18        |
| 2.2.1.5.1. Etki Karışımı ve Etkileşim.....                             | 19        |
| 2.2.1.5.2. Çoklu Bağlantı Sorunu.....                                  | 19        |
| 2.2.1.5.3. Sıfır Frekans ve Tam Ayırma Sorunu.....                     | 20        |
| 2.2.1.6. Model Seçim Stratejisi .....                                  | 20        |
| 2.2.1.6.1. Tam Regresyon Yöntemi .....                                 | 21        |
| 2.2.1.6.2. İleriye Yönelik Seçim Yöntemi .....                         | 21        |
| 2.2.1.6.3. Geriye Dönük Eleme Yöntemi .....                            | 21        |
| 2.2.1.6.4. Adım-Adım Regresyon.....                                    | 21        |
| 2.2.1.7. Lojistik Regresyonda Açıklayıcılık Katsayıları ( $R^2$ )..... | 22        |
| 2.2.1.8. Model Uyum İyiliğinin Değerlendirilmesi.....                  | 23        |
| 2.2.1.8.1. Pearson ki-kare ve Sapma (Deviance) İstatistikleri.....     | 23        |
| 2.2.1.8.2. Hosmer ve Lemeshow Testi .....                              | 24        |
| 2.2.1.8.3. Sınıflandırma Tablosu.....                                  | 24        |
| 2.2.1.9. Model Yeterliliği İle İlgili Ölçüler.....                     | 25        |
| 2.2.1.9.1. Artıkların İncelenmesi .....                                | 25        |
| 2.2.1.9.2. Etkili Gözlemlerin Belirlenmesi .....                       | 25        |
| 2.2.1.10. Katsayıların Yorumlanması .....                              | 26        |
| 2.2.2. Çok Kategorili (Multinomial) Lojistik Regresyon Analizi .....   | 29        |
| 2.2.3. Sıralı (Ordinal) Lojistik Regresyon Analizi .....               | 30        |
| <b>3. GEREÇ VE YÖNTEM.....</b>                                         | <b>32</b> |
| 3.1. Araştırmanın Modeli.....                                          | 32        |
| 3.2. Araştırmanın Yapıldığı Grup .....                                 | 32        |
| 3.3. Verilerin Toplanması .....                                        | 32        |
| 3.4. Verilerin Analizi .....                                           | 34        |
| <b>4. BULGULAR.....</b>                                                | <b>36</b> |
| <b>5. TARTIŞMA .....</b>                                               | <b>69</b> |
| <b>SONUÇ ve ÖNERİLER .....</b>                                         | <b>78</b> |
| <b>KAYNAKLAR .....</b>                                                 | <b>81</b> |
| <b>ÖZGEÇMİŞ.....</b>                                                   | <b>90</b> |

## TABLolar DİZİNİ

|                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------|----|
| <b>Tablo 2.1.</b> RR İçin Örnek Tablo.....                                                                         | 9  |
| <b>Tablo 2.2.</b> OR için Örnek Tablo.....                                                                         | 11 |
| <b>Tablo 2.3.</b> Örnek Sınıflandırma Tablosu (Alpar, 2013).....                                                   | 25 |
| <b>Tablo 2.4.</b> Artık Değerler ve Uzaklıklar (Field, 2009).....                                                  | 26 |
| <b>Tablo 2.5.</b> Bağımsız Değişken İki Kategorili Olduğunda Lojistik Regresyon Modelinin Olasılık Değerleri ..... | 27 |
| <b>Tablo 4.1.</b> Yapılan Analize İlişkin Gözlem Sayısı Bilgileri.....                                             | 40 |
| <b>Tablo 4.2.</b> Bağımlı Değişken Kodlama Sembolü Bilgisi .....                                                   | 40 |
| <b>Tablo 4.3.</b> Modele Alınan Kategorik Değişkenlerin Sıklık ve Kodlama Bilgileri....                            | 41 |
| <b>Tablo 4.4.</b> Gözlenen ve Tahmin Edilen Sınıflandırma (Çapraz Geçerlilik) Tablosu                              | 42 |
| <b>Tablo 4.5.</b> Başlangıç Aşaması Model Bilgileri .....                                                          | 43 |
| <b>Tablo 4.6.</b> Modele Eklenmeyen Değişkenlerin Skor Bilgileri .....                                             | 43 |
| <b>Tablo 4.7.</b> Modele İlişkin Omnibus Testi.....                                                                | 44 |
| <b>Tablo 4.8.</b> Model Açıklayıcılık Değerleri .....                                                              | 45 |
| <b>Tablo 4.9.</b> Hosmer & Lemeshow Model Uyum İyiliği .....                                                       | 45 |
| <b>Tablo 4.10.</b> Sınıflandırma Tablosu.....                                                                      | 45 |
| <b>Tablo 4.11.</b> Doymuş Model Değişkenleri ve Anlamlılık Değerleri .....                                         | 47 |
| <b>Tablo 4.11.</b> Doymuş Model Değişkenleri ve Anlamlılık Değerleri (devam).....                                  | 48 |
| <b>Tablo 4.12.</b> Çoklu Bağlantı Problemi Göstergeleri .....                                                      | 49 |
| <b>Tablo 4.13.</b> İleriye Dönük Modele İlişkin Omnibus Testi Sonuçları .....                                      | 50 |
| <b>Tablo 4.14.</b> İleriye Dönük Model İçin Uyum İyiliği Ölçütleri.....                                            | 51 |
| <b>Tablo 4.15.</b> İleriye Dönük Model İçin Hosmer & Lemeshow Testi Sonuçları .....                                | 51 |
| <b>Tablo 4.16.</b> İleriye Dönük Model İçin Sınıflandırma Tablosu.....                                             | 52 |
| <b>Tablo 4.17.</b> İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri .....                               | 53 |
| <b>Tablo 4.17.</b> İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri(devam)                              | 54 |
| <b>Tablo 4.17.</b> İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri(devam)                              | 55 |
| <b>Tablo 4.17.</b> İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri(devam)                              | 56 |
| <b>Tablo 4.18.</b> İleriye Dönük Modelden Anlamlı Değişkenler Çıkarıldığında Elde Edilen Değişimler.....           | 57 |
| <b>Tablo 4.19.</b> Modele Dâhil Edilmeyen Değişkenler .....                                                        | 58 |
| <b>Tablo 4.20.</b> Etkin Gözlemlerin Çıkarılmasıyla Elde Edilen Model .....                                        | 59 |
| <b>Tablo 4.20.</b> Etkin Gözlemlerin Çıkarılmasıyla Elde Edilen Model (devam).....                                 | 60 |
| <b>Tablo 4.21.</b> Çalışmaya Alınan Müşterilerin Bölgelere Göre Genel Özellikleri .....                            | 61 |
| <b>Tablo 4.22.</b> Çok Kategorili Lojistik Modele İlişkin Uyum Bilgileri.....                                      | 64 |
| <b>Tablo 4.23.</b> Çok Kategorili Lojistik Modele İlişkin Uyum İyiliği Sonuçları.....                              | 64 |
| <b>Tablo 4.24.</b> Çok Kategorili Modele İlişkin $R^2$ Değerleri .....                                             | 64 |
| <b>Tablo 4.25.</b> Olabilirlik Oran Testi Sonuçları.....                                                           | 64 |
| <b>Tablo 4.25.</b> Olabilirlik Oran Testi Sonuçları (devam) .....                                                  | 65 |
| <b>Tablo 4.26.</b> Modele İlişkin Katsayı Tahminleri ve Anlamlılık Düzeyleri .....                                 | 66 |
| <b>Tablo 4.26.</b> Modele İlişkin Katsayı Tahminleri ve Anlamlılık Düzeyleri (devam) ..                            | 67 |
| <b>Tablo 4.27.</b> Model Başarı Oranını Gösteren Sınıflandırma Tablosu .....                                       | 67 |

## ŞEKİLLER DİZİNİ

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| Şekil 2.1. Lojistik Fonksiyonu .....                                                            | 13 |
| Şekil 4.1. Çalışmaya Katılan Hastaların Cinsiyet Oranları.....                                  | 37 |
| Şekil 4.2. Çalışmaya Katılan Hastaların Başvuru Şikâyetleri.....                                | 37 |
| Şekil 4.3. Çalışmaya Alınan Hastaların Kalp Krizi Risk Sınıfları.....                           | 38 |
| Şekil 4.4. Sağlık Sigortası Ödemesi Yapan Müşterilerin Cinsiyet Oranları.....                   | 61 |
| Şekil 4.5. Sağlık Sigortası Ödemesi Yapan Müşterilerin Yaşadıkları Bölgelerin<br>Oranları ..... | 62 |
| Şekil 4.6. Sağlık Sigortası Ödemesi Yapan Müşterilerin Sigara Kullanım Oranları.                | 62 |



## SİMGELER ve KISALTMALAR DİZİNİ

|                                 |                                                               |
|---------------------------------|---------------------------------------------------------------|
| <b>RR</b>                       | : Relative Risk                                               |
| <b>OR</b>                       | : Odds Ratio                                                  |
| <b>Z</b>                        | : İstenen Güven Düzeyinde Çift Yönlü Standart Normal z Değeri |
| <b><math>I_{E+}</math></b>      | : Etkenin Olduğu Grupta İnsidans Hızı                         |
| <b><math>I_{E-}</math></b>      | : Etkenin Olmadığı Grupta İnsidans Hızı                       |
| <b><math>\beta_0</math></b>     | : Bağımsız Değişken                                           |
| <b><math>\beta_1</math></b>     | : Bağımsız Değişkenin Regresyon Katsayısı                     |
| <b><math>\varepsilon</math></b> | : Hata Terimi                                                 |
| <b>LRA</b>                      | : Lojistik Regresyon Analizi                                  |
| <b>LR</b>                       | : Lojistik Regresyon                                          |
| <b>X</b>                        | : Bağımsız Değişken                                           |
| <b><math>\pi</math></b>         | : Olasılık                                                    |
| <b>YAEKK</b>                    | : Yeniden Ağırlıklandırılmış En Küçük Kareler                 |
| <b>EKK</b>                      | : En Küçük Kareler                                            |
| <b>j</b>                        | : Grup Sayısı                                                 |
| <b><math>n_j</math></b>         | : Deneme                                                      |
| <b><math>r_j</math></b>         | : Başarı Sayısı                                               |
| <b>W</b>                        | : Wald Test İstatistiği                                       |
| <b>AIC</b>                      | : Akaike Information Criterion                                |
| <b>BIC</b>                      | : Bayesian Information Criterion                              |
| <b><math>R^2</math></b>         | : Açıklayıcılık Katsayıları                                   |
| <b>n</b>                        | : Gözlem Sayısı                                               |

|             |                                                          |
|-------------|----------------------------------------------------------|
| $y_i$       | : Bağımlı Değişken                                       |
| $\bar{y}$   | : Bağımlı Değişkenin $y$ 'nin Ortalaması                 |
| $\hat{y}_i$ | : Model Yardımıyla Kestirilen Değer                      |
| $M_\alpha$  | : Sabit Terimli Modelin Log-Olabilirliği                 |
| $M_\beta$   | : Bağımsız Değişkenlerin Olduğu Modelin Log-Olabilirliği |
| SS          | : Standart Sapma                                         |
| Maks        | : Maksimum                                               |
| Egz         | : Egzersiz                                               |
| EBA         | : Egzersize Bağlı Anjin                                  |
| -2LL        | : -2log likelihood                                       |
| ABD         | : Amerika Birleşik Devletleri                            |
| AKŞ         | : Açlık Kan Şekeri                                       |
| EKG         | : Elektrokardiyografi                                    |
| VKİ         | : Vücut Kitle İndeksi                                    |
| ANOVA       | : Tek Yönlü Varyans Analizi                              |
| AP          | : Anjina Pektoris                                        |
| MI          | : Myokard Infarktüsü                                     |
| VIF         | : Variance Inflation Factor                              |
| Exp         | : Exponential                                            |
| USD         | : United States Dollar                                   |
| HKO         | : Hata Kareler Ortalaması                                |

# 1. GİRİŞ

Teknolojinin gelişmesi ile birlikte daha büyük veriler depolanıp bu veriler üzerinde analizler yapılabilmektedir. Günümüzde birçok alanda yapılan çalışmalarda açıklanmak istenilen konularda kullanabilecek değişkenler fazla sayıda olabilir. Bu değişkenlerin birbiri ile arasındaki ilişkiler incelenebilir ve bu değişkenler tek tek veya birlikte olduğu zaman bağımlı değişken üzerindeki etkilerinin değişkenliği öğrenilmek istenebilir. Araştırmacılar, çalışmalarını daha hızlı, düzenli ve anlaşılabilir olması için bilgisayar programlarından yararlanarak elde etmiş olduğu verilerden bir model oluşturup bu model üzerinden sonuca ulaşabilirler.

Değişkenlerin fazla olduğu çok değişkenli istatistiksel analizlerde değişik analizler uygulanabilmektedir. Burada önemli olan bağımlı (dependent) değişkenin özelliklerine göre uygun olan istatistiksel analizi belirlemektir.

Bu çalışmada bağımlı değişkenin sayısal veya kategorik olduğu durumlarda kullanabilecek lojistik regresyon analizi (LRA) incelenmiştir. Bu amaçla açık erişime sahip iki farklı veri kümesi kullanılmıştır.

1845’li yıllara dayanan LRA ilk önceleri toplumdaki nüfus artışını, matematiksel bir ifadeyle açıklanmak için yapılan araştırmalarda ortaya çıkmıştır (1). Berkson tarafından 1944, 1953 ve 1955 yıllarında LR çalışmaları yapılmış ve bu çalışmalar ışığında LR’nin probit analizine bir alternatif olacağını 1972 yılında Finney öngörmüştür (2).

İstatistik paket programlarının kullanımının artmasıyla özellikle 2000’li yılların başından bu yana sağlık alanı, ekonomik tahminler, askeri konular, meteoroloji, iç göç hareketleri ve eğitim alanı başta olmak üzere diğer birçok alanda LRA’nın kullanımının arttığı görülmektedir (3–5).

Çalışmanın ikinci bölümünde lojistik regresyon analizine ilişkin genel literatür bilgileri verilmiştir. İlk olarak LRA’nın kullanım alanlarından, tercih nedenlerinden bahsedilmiştir. Daha sonra analize katılacak değişkenlerin kodlanması, risk kestirimleri ve lojistik regresyon analizinin çeşitleri detaylı bir şekilde anlatılmıştır.

İki kategorili lojistik regresyon analizinde kullanılan kalp krizi verisinde hastaların yaşı, dinlenme anı kan basıncı, maksimum nabız sayısı, ST segmenti çökme değeri ve görüntülenen damar sayısı, düşük ve yüksek MI risk sınıfları arasında anlamlı farklılık göstermiştir ( $p<0,001$ ).

Çok kategorili lojistik regresyon analizinde kullanılan sağlık sigortası verisinde, ABD müşterilerin yaşadıkları bölgeler bağımlı değişken olarak alınarak diğer değişkenlerin yaşanılan bölge üzerindeki etkileri belirlenmeye çalışılmıştır. Müşterilerin yaşları, cinsiyetleri, sigara kullanımı ve çocuk sayıları bölgeler arasında anlamlı farklılığa sahip bulunmamıştır ( $p>0,05$ ). Ancak, müşterilerin VKİ değerleri güneydoğu eyaletlerinde yaşayanlar için anlamlı düzeyde yüksektir ( $p<0,001$ ). Dolayısıyla müşterilerin ödediği sağlık sigorta miktarı da güneydoğu eyaletlerindeki müşterilerinde anlamlı düzeyde yüksek bulunmuştur ( $p=0,010$ ).

## 2. GENEL BİLGİLER

İstatistik, bilimsel araştırma ve çalışmalarda neredeyse bütün bilim dallarında kullanılan ve hemen hemen her konuda uygulama alanı bulan bir karar verme yöntemidir (6).

Özarıcı'ya göre istatistiğin amacı, “deneysel çalışmalar ve örnekleme nedeniyle ortaya çıkan belirsizlik ve noksan bilginin akla ve belirli ilkelere uygun olarak tartışılacağı bir anlatım biçimi geliştirmek ve bu anlatım biçimine uygun olarak düşünülüp kararlaştırılmış çıkarsama veya kararlara ulaşmanın araçlarını bulup ortaya koymaktır. Kullanılan bir çok teknikle çok sayıda değişkenin oluşturduğu sistemin yapısı belirlenir ve olabildiğince basit bir forma dönüştürülerek herhangi bir konuda karar için gerekli bilgi çıkartılır. Araştırmacı amacını gerçekleştirirken kullanacağı istatistiksel yöntem ve teknikleri bağımlı-bağımsız varsa etkisi arındırılacak değişkenlerin varlığına, bu değişkenlerin sayısına, ölçek türüne vb. göre belirleyecektir. Kısaca verilerin niteliklerine göre kullanılacak istatistiksel yöntem ve teknikler değişiklik gösterecektir” (7).

Birçok değişken birbirinden etkilenir ve birisi diğerinin nedeni olabilir. Değişkenler bazı durumlardan olumlu ve olumsuz etkilenip bu etkilenmeler yüksek düzeyde veya düşük düzeyde olabilir. Bahsedilen nedeni belirleme ve etki düzeylerini belirlemede çok değişkenli istatistiksel yöntemler kullanılır (8).

İstatistiksel bir analiz olan regresyon metodu ise iki veya daha fazla değişkenin aralarındaki sebep-sonuç ilişkisi ile ilgili çıkarımlar yapabilmek için matematiksel bir model kurularak oluşturan bu ilişkinin analiz edilmesi tekniğine verilen isimdir. Bir örnekten elde edilen tahminin geçerliliği, tahminin elde edilirken kullanılan metodun geçerliliğiyle doğrudan ilişkili olduğundan regresyon analizinde parametre tahmininde bulunulmasının yanı sıra bu tahminlerde kullanılan yöntem veya tahmin ediciler üzerinde durulur (9).

Genel olarak regresyon analizi, bağımlı ve bağımsız değişkenlerin aralarındaki neden-sonuç ilişkisini istatistiksel bir model ile açıklamaktadır. Bağımlı değişken değerinin kestirimi bu varsayılan ilişkiye dayanarak yapılır. Genelde mevcut bir popülasyondan alınan örneklemelerden elde edilir (10). Sigara tüketiminin eğitim, gelir,

yaş ve sigara fiyatı gibi çeşitli demografik ve sosyoekonomik özelliklerle ilişkisi olup olmadığı incelenmek istenirse ya da bir gayrimenkul değerlendirme uzmanı, bir evin satış fiyatını binanın seçilen fiziksel özellikleri ile binaya ödenen vergiler üzerinden ilişkilendirmek isterse değişkenler arasındaki işlevsel ilişkiyi araştırmak için regresyon yöntemini kullanabilir.

Regresyon yöntemi bağımsız değişkenlerin sayılarına göre farklı isimlendirilmektedir. Elimizde bir bağımsız değişken olduğu zaman tek değişkenli, iki ve daha fazla değişken olduğu zaman çok değişkenli regresyon analizi olarak adlandırılır.

Bilinen doğrusal regresyon analizinde kullanılan bağımlı ve bağımsız değişkenler sayısal olarak belirtilir. Lojistik regresyonda bağımlı değişkenin iki seviyesi vardır (kategorik) ve bu değerler 0 ve 1 olarak yer alırlar. Bağımlı değişkenin 2'den fazla değeri (seviyesi) olması durumunda çok terimli (multinomial) lojistik regresyon halini alır.

## **2.1. Lojistik Regresyon Analizi**

Lojistik regresyon analizi, bağımlı değişken iki ya da daha fazla kategoride olduğunda, bağımsız değişkeninde hem sayısal verilerden hem de kategorik verilerden oluştuğu zaman en iyi modeli kurarak risk faktörlerinin belirlenmesinde kullanılan bir analiz yöntemidir.

Kategorik verilerin çok değişkenli istatistiksel analizi, neredeyse tüm alanlarda kullanılmaktadır. Çeşitli analizlere kıyasla çok daha elverişli ve regresyon tipi çalışma mantığıyla avantaj sağlayan LRA kategorik verilerin analiz edilmesinde de önemli bir yere sahiptir (11).

LRA 3 farklı kategoride yapılabilmektedir. Yapılacak olan analizin lojistik regresyon analizinin hangi kategorisine girdiğini belirlemek için bağımlı (sonuç) değişkenin özelliğine bakılabilir. Eğer bağımlı değişkenimiz;

- iki kategorili niteliksel bir değişken ise iki kategorili (binary),
- ikiden fazla sıralı olmayan niteliksel bir değişkene çok kategorili (multinomial),

- ikiden fazla sıralanabilir olduğu durumlarda çok kategorili sıralanabilir (ordinal) lojistik regresyon yöntemleri kullanılır.

Bu yöntemlerin dışında koşullu (conditional) lojistik regresyon ve kesin (exact) lojistik regresyon yöntemleri de vardır. Koşullu lojistik regresyon yöntemi çoğunlukla sağlık alanında kullanılan olgu-kontrol çalışmalarında kontrol grubu seçimi için, kesin lojistik regresyon yöntemi ise gözlem sayımızın yani elimizdeki verilerin az olduğu zamanlarda kullanılır.

Kategorik verilerin en sık karşılaştığımız biçimi var-yok veya başarı-başarısız gibi iki cevaplı sonuçlanan dikatom verilerdir (dichotomous). Beş cevaplı olarak cevaplanan sonuçlarda vardır. Örneğin Likert tipi skala sorular için yok-biraz var-orta-fazla var-çok var veya ters yönlü ölçek sorularında olduğu gibi. Tabii ki sonuçların meydana gelmesinde farklı sebepler vardır. Acaba faktörlerin farklılaşmaları ve değişik birleşmeleri, neticenin ortaya çıkması veya çıkmamasında, oluşumun derecelendirilmesinde nasıl bir katkısı vardır? Normal dağılımın olmadığı hallerde sonucun meydana gelmesi-gelmemesi, düşük-orta-şiddetli olarak derecelendirilmesinde açıklayıcı değişkenlerin etkileri nasıl meydana çıkarılabilir? Birçok etken içinde hangi etken ya da etkenler, ne seviyede bu sonuçların meydana gelmesine zemin hazırlamaktadır? Bu etkenlerin hangisinin daha önemli risk faktörü olduğu nasıl bulunabilir gibi sorulara yanıt almak adına lojistik regresyon analizinden faydalanılabilir (12).

### **2.1.1. Lojistik Regresyon Analizi Tanımı ve Tarihçesi**

Lojistik veya logit regresyon; açıklayıcı değişkenlere göre ikili hedef değişkeninden ortaya çıkan sonuçların olasılık olarak ulaşıldığı bir regresyon yöntemidir (13). Bu yöntem ile elde edilen lojistik regresyon modeli, bağımsız değişkenlerle, bağımlı değişken arasındaki ilişkiyi tahmin etmede kullanılır (14). Lojistik regresyon tahmin edici (predictive) bir analiz yöntemidir.

Bağımlı değişkenin iki veya ikiden fazla kategorik veri olduğu durumlarda bağımsız değişkenlerle sebep-sonuç ilişkisini belirlemede kullanılan LRA 1844-1845’li yıllara dayanmaktadır. Lojistik model sosyo-ekonomik ağırlıklı konuları incelemesinin yanı sıra Pierre François Verhulsts tarafından ilk önceleri toplumdaki

nüfus artışını, matematiksel bir ifadeyle açıklanmak için yapılan araştırmalarda ortaya çıkmıştır (1).

Berkson tarafından 1944, 1953 ve 1955 yıllarında lojistik regresyon çalışmaları yapılmış ve bu çalışmalar ışığında lojistik regresyonun probit analizine bir alternatif olacağını 1972 yılında Finney öngörmüştür. 1967 ve 1971 yılında da lojistik regresyonun, normal dağılım varsayımı olan varyansın 1 ve ortalamasının 0 olması görmezden gelindiğinde diskriminant analizine alternatif olacağı söylenmiştir. Eklemeli olasılık modellerindeki etkileşimi ortadan kaldırmak için lojistik regresyon 1975 yılında Koch tarafından önerilmiştir (15).

### **2.1.2. Lojistik Regresyon Analizinin Tercih Nedenleri**

En bilinen regresyon yöntemlerinden olan basit ve çoklu doğrusal regresyon bağımlı ve bağımsız değişken arasındaki ilişkiyi değişik amaçlarla inceleyen yöntemleridir. Basit ve çoklu doğrusal regresyonda bağımlı değişken sayısal veri türündedir. Araştırma konularına ait veriler her zaman sayısal bir veri olmayabilir. LRA, sayısal olmayan veriler karşımıza kategorik veri olarak çıktığı zaman verilerin analizinde kullanılan bir çok istatistiksel yöntemlerden birisidir (13).

Lojistik regresyon analizinin seçilmesinin önemli sebeplerinden ilki, fonksiyonunun 0 ve 1 değerleri arasında olmasıdır. Lojistik model meydana gelecek riski 0 ile 1 arasında tahminler. Yani risk 1'in üzerine veya 0'ın altına inemez (16).

2000'li yıllardan bu yana daha yaygın kullanılmaya başlanılan lojistik regresyon analizinin neden daha fazla tercih edildiği aşağıda belirtilen maddelerde açıklanmaktadır.

1. Lojistik regresyon bağımsız değişkenin sayısal ya da kesikli olmasına engel teşkil etmez.
2. Lojistik modelin parametreleri rahat değerlendirilebilir ve matematiksel olarak kullanımı anlaşılabilir rahatlıkta olan fonksiyonlar üretmektedir.
3. Lojistik modele bağlı analizleri rahatlıkla yapabilmeye yarayan çok sayıda istatistik programları vardır (SPSS, R, SAS, Stata vb.).

4. Bağımsız değişkenlerin olasılık fonksiyonlarının dağılımıyla alakalı bir sınırlaması yoktur.

5. Lojistik regresyonun tüm olasılıkları 0 ile 1 arasında değerler alır.

6. Bağımsız ve bağımlı değişken arasındaki fonksiyonun doğrusal olması şartı yoktur; doğrusal ilişkinin yanı sıra üstel veya polinom ilişkinin olması sorun teşkil etmez. Bunun sebebi de değişkenler arasında logit bir ilişki olduğunu farz edip doğrusal olmayan modeller ortaya çıkarabilir. Yani lojistik regresyon, doğrusal olmayan ilişkiyi bozmadan, logaritmik dönüştürmeler yaparak ilişkinin şeklini doğrusal hale döndürür (17).

### **2.1.3. Lojistik Regresyonun Kullanım Alanları**

İstatistik paket programlarının kullanımının artmasıyla özellikle 2000'li yılların başından bu yana sağlık alanı, ekonomik tahminler, askeri konular, meteoroloji, iç göç hareketleri ve eğitim alanı başta olmak üzere diğer birçok alanda lojistik regresyon analizinin kullanımının arttığı görülmektedir (3–5).

Bir hücrenin kanserli bir hücre mi yoksa kanserli olmayan bir hücre mi olduğunu anlamamıza yardımcı olacak bir model ortaya koymak, tedavinin hastalık üzerinde işe yarayıp yaramadığını anlamak gibi çeşitli amaçlar doğrultusunda oldukça sık kullanılmaktadır (18).

Toplumda yaşayan bireylerin çoğu özellikleri birbirlerine benzerken bazı hastalıklar bir grup insanda görülüp diğer insanlarda görülmeyebiliyor. Bu hastalığa sebep olan en önemli ya da diğerlerine göre önemsiz olan nedenlerin ayrımını yapabilmek için lojistik regresyon yönteminden yararlanılmaktadır (12).

### **2.1.4. Bağımsız Değişkenler**

Sağlık alanında yapılan çoklu regresyon modeli ile alakalı uygulamalarda bağımsız değişkenleri (predictors) genel olarak bir hastalığın olup olmamasını etkileyen değişkenler ya da risk değişkenleri olarak adlandırılır. Sigara içmek bir hastalığın ortaya çıkıp çıkmamasını etkileyen bir durum ise sigara içme değişkeni risk değişkeni ya da riski oluşturan değişkendir. Lojistik regresyon analizinde risk

değişkeni niteliksel ya da niceliksel veri türünde olabilmektedir (13).

### 2.1.5. Bağımlı Değişkenlerin Kodlanması

İki kategorili lojistik regresyon analizinde bağımlı veya cevap değişkenin (outcome) çoğunlukla riskin olduğu hallerde: 1, riskin olmadığı hallerde: 0 ile kodlanması daha faydalıdır. Başarılı olma: 1, başarısız olma: 0 ile göstermekte yarar vardır. Yani genel olarak bir olayın meydana gelmesi: 1, gelmemesi: 0 ile kodlanabilmektedir. Çok kategorili durumlarda bu sayılar kodlamalar kategori sayısına göre artmaktadır (13).

### 2.1.6. Risk Kestirimi

Lojistik regresyon sonuçlarının yorumlanmasında gereksinim duyulacak risk kavramı ve risk kestirimine ilişkin bazı temel ölçülerin bilinmesi gerekir. Bu ölçüler genel risk ölçüleri ve göreceli risk ölçüleri olarak ele alınırlar.

#### 2.1.6.1. Genel Risk Ölçüleri

Popülasyondaki hastalık düzeyini belirleyen genel risk ölçüleri insidans ve prevalanstır. Bunlar hastalığa neden olan risk faktörlerini dikkate almayıp genel riski belirledikleri için çok fazla dikkate alınmazlar (13).

Sağlık Bakanlığının yapmış olduğu tanıma göre insidans “*risk altındaki sağlam kişilerin belirli sürede, belirli bir hastalığa yakalanma olasılığını gösteren ölçüttür*” (19). Diğer bir deyişle, incelenen sürede yeni ortaya çıkan olgularla ilgilenilir. Bu sebeple insidans hızına atak hızı da denir (13). İnsidans hızı Eşitlik (1) ile verilir.

$$\text{İnsidans Hızı} = \text{Yeni Olgu Sayısı} / \text{Risk Altındaki Nüfus} \quad (1)$$

Yine Sağlık Bakanlığının yapmış olduğu tanıma göre prevalans “*Belirli bir süre içinde bir hastalığın toplumda görülme sıklığını gösteren ölçüttür*” (19).

İncelenen süre içinde eskiden başlayan olgular da dikkate alınır (13). Prevalans hızı Eşitlik (2) ile verilir.

$$\text{Prevalans Hızı} = (\text{Eski} + \text{Yeni Olgu Sayısı}) / \text{Risk Altındaki Nüfus} \quad (2)$$

### 2.1.6.2. Bağımlı Risk Ölçüleri

Bağımlı Risk (Relative Risk-RR) ve Odds Oranı (Odds Ratio-OR) bir olgunun ortaya çıkışındaki etkenleri göz önüne alan iki önemli risk ölçüsüdür. Risk kestirimi için kullanılan ölçüler yapılan çalışmanın tipine bağlı olarak değişmektedir. İleriye yönelik çalışmalarda insidans hızlarından yararlanılarak RR hesaplanabilirken geriye dönük çalışmalarda (olgu-kontrol) insidans hızları hesaplanamadığından RR hesaplamakta olası değildir. Geriye dönük çalışmalarda olası olmayan bu durum için risk kestirimi OR hesaplanarak bulunur. Çalışmanın konusu olan lojistik regresyon yönteminde de risk kestirimi için OR hesaplanmaktadır (13).

### İleriye Yönelik Araştırmalar (Prospective Studies) Ve Bağımlı Risk Kestirimi

**Tablo 2.1.** RR İçin Örnek Tablo

| Etken  | Hastalık |       | Toplam    |
|--------|----------|-------|-----------|
|        | Var      | Yok   |           |
| Var    | a        | b     | $a+b=n_1$ |
| Yok    | c        | d     | $c+d=n_2$ |
| Toplam | $a+c$    | $b+d$ | n         |

$$RR = (\text{Etkenin Olduğu Grubun Riski}) / (\text{Etkenin Olmadığı Grubun Riski})$$

$$RR = (\text{Etkenin olduğu grupta İnsidans Hızı}) / (\text{Etkenin Olmadığı Grupta İnsidans Hızı})$$

$$\text{Etkenin olduğu grupta İnsidans Hızı: } I_{E+} = a/(a+b)$$

$$\text{Etkenin Olmadığı Grupta İnsidans Hızı: } I_{E-} = c/(c+d)$$

Bu bilgiler çerçevesinde görel risk Eşitlik (3) ile verilir.

$$RR = \frac{I_{E+}}{I_{E-}} = \frac{a/(a+b)}{c/(c+d)} \quad (3)$$

RR<1 ise etkenin hastalık riskini azaltıcı etkisi olduğu,

RR>1 ise etkenin hastalık riskini artırıcı etkisi olduğu söylenir.

### RR'in Standart Hatası ve Güven Aralıkları

RR'in standart hatasını bilmek güven aralıklarını bulmak ve hipotezlerimizi test edebilmek için gereklidir. RR'in örneklem dağılımının normal dağılım göstermesi için ln(RR)'si alınır. ln(RR)'nin standart hatası Eşitlik (4) ile verilir.

$$SH(\ln(RR)) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}} \quad (4)$$

Buradan, RR için güven aralığı Eşitlik (5) ile elde edilir.

$$e^{\ln(RR) \pm z_{1-\alpha/2} \times SH(\ln(RR))} \quad (5)$$

z; istenen güven düzeyinde çift yönlü standart normal z değeridir.

Güven aralığının “1” i içermesi durumunda elde edilen RR'nin istatistiksel açıdan önemsiz olduğu söylenir (13) .

### Geriye Dönük (Retrospektif) Araştırmalar ve OR Kestirimi

OR geriye dönük olgu-kontrol çalışmalarında kullanılan bir risk ölçüsüdür. OR'yi hesaplamak için Tablo 2.2'deki gibi dört gözlü tablodan yararlanılır.

**Tablo 2.2.** OR için Örnek Tablo

| Etken  | Hastalık           |                    | Toplam |
|--------|--------------------|--------------------|--------|
|        | Var                | Yok                |        |
| Var    | a                  | b                  | a+b    |
| Yok    | c                  | d                  | c+d    |
| Toplam | a+c=n <sub>1</sub> | b+d=n <sub>2</sub> | n      |

OR, herhangi bir olayın olma olasılığının (p), olmama olasılığına (1-p) oranı olarak tanımlanır (Eşitlik 6). Bazı kaynaklarda olayın olma olasılığı  $\pi$ , olmama olasılığı  $(1 - \pi)$  ile gösterilmektedir.

$$\text{Odds} = p / (1 - p) = \pi / (1 - \pi) \quad (6)$$

Eşitlik (6) ve Tablo 2.2’den,

Hasta grupta odds: a / c

Sağlıklı grupta odds: b / d olarak tanımlanır.

İki odds ’un oranına OR denir ve Eşitlik (7) ile verilir.

$$\text{OR} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c} \quad (7)$$

OR<1 ise etkenin hastalık riskini azaltıcı etkisi olduğu,

OR>1 ise etkenin hastalık riskini artırıcı etkisi olduğu söylenir.

### **OR’in Standart Hatası ve Güven Aralıkları**

ln(OR)’nin standart hatası ve güven aralığı sırasıyla Eşitlik (8) ve (9) ile verilir.

$$SH(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad (8)$$

$$e^{\ln(OR) \pm Z_{1-\alpha/2} \times SH(\ln(OR))} \quad (9)$$

## 2.2. Lojistik Regresyon Çeşitleri

### 2.2.1. İki Kategorili (Binary) Lojistik Regresyon

Bağımlı değişkenin iki kategorili olduğu durumlar (evet-hayır, var-yok) için kullanılır (12). Daha öncede bahsedildiği gibi genel olarak riskin olmadığı durum için 0, olduğu durum için 1 kodlanır.

Bağımsız değişkene ilişkin veri matrisi  $X$ 'in değeri bilindiğinde olayın olma ihtimali  $\pi$  ile gösterilsin. Standart doğrusal model kullanılarak  $\pi$ 'yi tanımlamak Eşitlik (10)'daki gibi olur.

$$\pi(x) = P(Y=1 / X=x) = \beta_0 + \beta_1 x + \varepsilon \quad (10)$$

$\beta_0$ : Bağımsız değişken

$\beta_1$ : Bağımsız değişkenin regresyon katsayısı

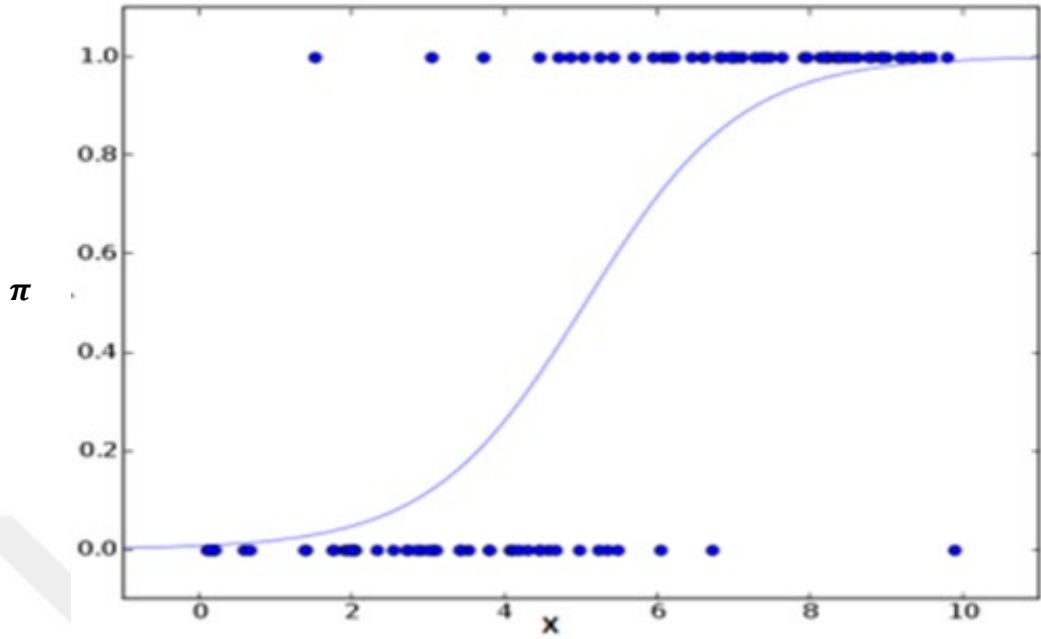
$\varepsilon$ : Hata terimi

Eşitlik (10)'da  $x$ 'in  $-\infty$  ve  $+\infty$  arasında değerler alabilmesinden dolayı,  $E(Y|x)$ 'in her değeri alabileceği görülmektedir. Bağımlı değişken iki kategorili olduğunda koşullu ortalama, sıfırla bir arasında değişmek zorundadır (20).

LRA'da bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak ifade edilir. Burada çoklu bağlantı sorununun olmaması şart koşulmasa da analiz bundan etkilenmektedir (21).

#### 2.2.1.1. Lojistik Regresyon Modeli

$X$  bağımsız değişkenleri ile olasılık  $\pi$  arasındaki ilişki, genel olarak lojistik yanıt fonksiyonu ( S biçiminde eğri) ile gösterilmektedir ( Şekil 2.1). İlk olarak  $X$  değerlerindeki yükselme ile beraber  $\pi$  değeri yavaşça artar, sonrasında artış hızlanır ve en sonunda durağanlaşır.  $\pi$  değeri hiçbir durumda 1'in üstüne çıkmaz. Buradan bağımsız değişkenler ile olasılık arasındaki ilişkinin doğrusal olmadığını çıkarabiliriz (13,22).



**Şekil 2.1.** Lojistik Fonksiyonu

Şekil 2.1’deki lojistik fonksiyonunu tanımlayan lojistik regresyon modeli Eşitlik (11) ile verilir. Bu lojistik regresyon modeli  $X=x$  olduğunda  $Y=1$  olması olasılığı  $\pi$ ’dir şeklinde yorumlanabilir.

$$\pi(x) = P(Y=1 / X=x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (11)$$

Eşitlik (11)’de verilen fonksiyonu doğrusallaştırabilmek için lojit dönüşüm uygulanmalıdır. Bunu odds’unun doğal logaritmasını alarak yapmak gerekir.

Odds oranının doğal logaritması alınarak doğrusal modele dönüştürülerek lojit dönüşümü yapılmış olur ve lojitler  $-\infty$  ile  $+\infty$  arasında değer alabilir (Eşitlik 12). Lojit, odds ’un doğal logaritmasıdır.

$$\text{lojit } \pi(x) = g(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) \quad (12)$$

$$\frac{\pi(x)}{1 - \pi(x)} = e^{(\beta_0 + \beta_1 x)} \quad (13)$$

Odds’un doğal logaritması alındığında model doğrusal modele dönüşür.

$$g(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \ln e^{(\beta_0 + \beta_1 x)} = \beta_0 + \beta_1 x \quad (14)$$

Tek deęişkenli model Eşitlik (13) veya Eşitlik (14)'teki gibi yazılır.

LR modelinin tek varsayımı lojit ile bağımsız deęişkenler arasındaki ilişkinin doğrusal olmasıdır.

Olasılık deęeri 0-1 arasında deęerler alırken,  $x'$  in deęerlerine baęlı olarak lojit  $-\infty$  ile  $+\infty$  arasında deęerler alabilmektedir. Lojistik model  $x'$ deki 1 birim deęişimin lojit'te ne kadarlık deęişime neden olduğunu gösterir (13,22,23).

Doğrusal regresyonda artıklar 0 ortalama ve sabit varyansla normal dağılırken, lojistik regresyonda artıklar 0 ortalama ve  $\pi(1 - \pi)$  varyansla binom dağılım gösterir (13).

#### **2.2.1.2. Regresyon Katsayılarının Kestirimi**

İki kategorili bir lojistik regresyon modelinin katsayılarının tahmininde; yeniden ağırlıklandırılmış en küçük kareler (YAEKK) yöntemi, en çok olabilirlik yöntemi, minimum lojit ki-kare yöntemi en çok tercih edilen yöntemlerdendir (24).

##### **2.2.1.2.1. En Çok Olabilirlik Yöntemi**

Doğrusal regresyonda parametrelerin tahmininde kullanılan en küçük kareler (EKK) yöntemi bağımsız deęişkenin normal dağılım varsayımını taşıdığı için lojistik regresyon analizinde yeniden ağırlıklandırılmış en küçük kareler yöntemi, en çok olabilirlik yöntemi, minimum lojit ki-kare yöntemi yaygın olarak kullanılır (24).

En çok olabilirlik yönteminde, örneklem büyüklüğü ( $n$ ) sonsuza yaklaştıkça yöntem tutarlıdır ve asimtotik olarak normal dağılır. Örneklem küçük olduğu zamanda kötü sonuçlar vermemekle birlikte deęişen sayısının en az 10 katı gözlem olması gerekmektedir (13).

Bu yöntemde amaç gözlediğimiz veri kümesini elde etme olasılığını en çok yapacak parametreleri bulmaktır. Onun için ilk olarak olasılık fonksiyonun oluşturulması gerekir. En çok olabilirlik tahminleri gözlenen deęerlere en yakın olacak şekilde yani fonksiyonu maksimum yapacak şekilde seçilir (24). Model seçimi yapıldıktan sonra doğruluğunu test etmek için Pearson ki-kare istatistięi ve sözde  $R^2$

(pseudo  $R^2$ ) değerlerine bakılır ve böyle bir durumda değeri yüksek olan sözde  $R^2$  hangi modelin analiz sonucunu daha iyi tahmin ettiğini belirler (25).

Üzerinde araştırma yapılan verilerde parametre sayısı denek sayısına göre fazla ise koşullu (conditional) en çok olabilirlik yöntemi, parametre sayısı denek sayısına göre az ise koşulsuz (unconditional) en çok olabilirlik yöntemi kullanılmalıdır. Koşullu en çok olabilirlik yöntemi yansız sonuçlar verirken koşulsuz yöntemin yanlı sonuçlar verdiği gözlenmiştir. Hangi yöntemin kullanılacağından tam emin olunmadığı zaman koşullu en çok olabilirlik yönteminin kullanılması önerilmektedir (26).

#### **2.2.1.2.2. Yeniden Ağırlıklandırılmış En Küçük Kareler Yöntemi**

EKK yöntemi doğrusal regresyonda sabit varyans (homoscedasticity) olduğu durumda kullanılırken lojistik regresyon yönteminde varyansın sabit olmadığı (heteroscedasticity) bilinmektedir. (YAEKK) yöntemi sabit varyanslılık durumu olmadığı zaman uygulanan bir yöntemdir. Gruplanmış verilerde  $j$  grup sayısı,  $n_j$  deneme,  $r_j$  başarı sayısı olsun. Bu durumda başarı oranı  $P_j = \frac{r_j}{n_j}$  olur. Buradan  $var(r_j/n_j) = p_j = (1 - p_j/n_j)$  olur.  $Lojit(r_j/n_j)$ 'nin bağımsız değişkenler üzerinde  $w_j = n_j/p_j (1 - p_j)$  ağırlığı ile ağırlıklandırılmış regresyonu yapılmalıdır.  $P_j$ 'nin bir fonksiyonu olan  $w_j$  ağırlık değeri EKK yönteminin iteratif olarak uygulanarak her adımda yeniden elde edilmesiyle bulunur (27,28).

#### **2.2.1.2.3. Minimum Lojit Ki-Kare Yöntemi**

Berkson tarafından geliştirilen minimum lojit ki-kare yöntemi 2xJ çapraz tablolarında beklenen ve gözlenen lojit değerleri arasındaki farkı kullanır. Tekrarlı verilerin olduğu zaman kullanılan bu yöntem YAEKK'nın özel bir biçimidir.

YAEKK yönteminde verilen  $P_j$  olasılığına yapılan lojit dönüşüm, minimum lojit ki-kare yönteminde yanıt değişkenini vermektedir ve kestirimde kullanılan ağırlık " $n_j/p_j (1 - p_j)$ " değerindedir. Burada en küçük kareler kestirimine ulaşılır ve bu ulaşılan EKK kestirimi minimum lojit ki-kare kestirimidir (29,30).

### 2.2.1.3. Katsayıların Önemliliğinin Test Edilmesi

Katsayıların tespit edilmesinden sonra bu katsayıların/değişkenlerin önemliliğinin test edilmesi gerekir. Modelde bir değişkenin bulunması, değişkenin bulunmamasından daha fazla bilgi veriyor mu? sorusuna cevap aranır. Değişkenlerin önemliliği; Olabilirlik Oran (likelihood ratio), Wald veya Skor testleriyle incelenebilir.

#### 2.2.1.3.1. Olabilirlik Oran Testi

Bağımsız değişkenin önemliliği, olabilirlik G istatistiği ile incelenir.

$$LR = G = -2\ln \left( \frac{L(\text{değişken modelde olmadığında})}{L(\text{değişken modelde olduğunda})} \right) \quad (15)$$

Parantezin içindeki olabilirlik istatistiği, sapma (deviance) olarak ifade edilir. Logaritmanın -2 katı alınarak dağılımı bilinen bir değere ulaşıp matematiksel olarak anlamlı bir değer elde edilebilmektedir. Bu teste, olabilirlik oran testi denilmektedir. Test işlemi modeli uydurmak için gözlem sayısı (n) yeterince büyük olduğunda geçerlidir. Ulaşılan test değerinin küçük olması, modele eklenen değişkenlerin lojit'in kestiriminde önemli bir katkı sağlamadığını ve değişkenlerin modelde bulunmasına gerek olmadığını gösterir (13,31,32).

#### 2.2.1.3.2. Wald Testi

Değişkenlerin önemliliğinin testinde kullanılan diğer bir yöntem Wald testidir. Wald testi,  $\beta_1$  in en çok olabilirlik tahminine ve bu tahmine ait standart hatanın karşılaştırılmasına dayanmaktadır. Wald test istatistiği standart normal dağılıma uyar. Test istatistiği Eşitlik (16)'daki gibidir.

$$W = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)} \quad (16)$$

Kovaryans matrisinin köşegen elemanlarının kareköklerinin alınmasıyla  $\beta_1$ 'in standart hatası elde edilmiş olur.  $\beta_1$ 'in en çok olabilirlik kestiriminin karesi standart hatasının karesine bölüldüğünde de Wald istatistiği ki-kare dağılımına uymaktadır Eşitlik (17).

$$W = \frac{\hat{\beta}_j^2}{s(\hat{\beta}_j)^2} \sim X^2 \quad (17)$$

Büyük örneklemelerde olabilirlik oran testi ve Wald testi benzer sonuçlar verirken küçük örneklemelerde hangi testin daha iyi sonuç verdiği hakkında literatür bilgisi çok olmamakla beraber Wald testi yerine olabilirlik oran testinin uygulanması belirtilmiştir (13,14).

#### 2.2.1.3.3. Skor Testi

Skor testinin en büyük avantajı en çok olabilirlik kestiriminin hesaplanmasına gereksinim duymadığı için hesaplama işlemlerini çok fazla kısaltır. P serbestlik dereceli ki-kare dağılımı gösteren bu istatistik denklemi Eşitlik (18)'deki gibidir (33).

$$ST = \frac{\sum_{i=1}^n x_i(y_i - \bar{y})}{\sqrt{y_i(1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (18)$$

Eşitlik (18)'deki score istatistiği LR'daki değişkene ait katsayının önemini test etmek için geçerli bir testtir (14). Önemi test edilen katsayıların odds oranları kullanılarak yorumlanması yapılır (34).

#### 2.2.1.4. Bağımsız Değişkenlerin Lojistik Regresyon Modeline Alınması

##### 2.2.1.4.1. Sayısal (Niceliksel) Değişkenler

Sayısal bağımsız değişkenler modele dâhil edilirken herhangi bir tanımlama işlemine gerek duyulmaz. Bu değişkenlerdeki bir birimlik değişimin lojitte ne kadarlık değişime neden olduğunu anlamak için sayısal değişkenlerin katsayılarına bakmamız gerekmektedir. Bağımsız değişken ile lojitin doğrusal olması varsayımı gerektiren bu durumda, Box-Tidwell testinden yararlanarak aralarındaki ilişki hakkında yorum yapılabilmektedir (13,35).

Sayısal bağımsız değişken x ile bu değişkene ait olan doğal logaritma  $\ln(x)$  çarpılır. Ortaya çıkan  $x\ln(x)$  etkileşim terimi modele eklenerek modelde önemli bulunması doğrultusunda lojit ile sayısal bağımsız değişken arasında doğrusal bir ilişkinin bulunmadığını bize gösterir (13,35).

#### **2.2.1.4.2. Niteliksel Değişkenler**

İki kategorili değişkenler modelde kullanılırken riskin düşük olduğu gruba daha az, riskin fazla olduğu gruba daha yüksek puan verilmesi yorumlama açısından kolaylık sağlar. İki kategorili değişkenlerde de istediğimiz kategoriye referans kategori olarak seçebiliriz (13).

Nominal değişkenleri kategorik değişkenler olarak tanımlarken kategori değerinin belirlenerek seçilmesi gerekir. SPSS programında Analyze>Regression>Binary Logistics sekmeleri tıklanarak açılan pencerede bağımlı ve bağımsız değişkenler atandıktan sonra niteliksel değişken için tanımlama Categorical komutu tıklanarak açılan pencerede Contrast'ın belirleyip bu kontrasta göre referans kategorinin belirlenmesiyle olur (13,36). Bu aşamada 7 farklı kontrast mevcuttur (13).

1. Indicatör: Referans kategori sıfır değeri alan kategoridir.
2. a) Simple/ Last: En büyük numaralı yani en son kategori referans kategoridir.  
b) Simple/First: En küçük numaralı yani ilk kategori referans kategoridir.
3. Difference: Kendinden önceki sınıfların ortalamasını referans kategori olarak alır.
4. Helmert: Kendinden sonraki sınıfların ortalamasını referans kategori olarak alır.
5. Repeated: Kendinden bir önceki sınıfı referans kategori olarak alır.
6. Deviation: Genel ortalamayı referans kategori yapar.
7. Polynomial: Sadece nicel değişkenler için olup kategoriler arasındaki riskteki değişimlerin doğrusal olup olmadığını inceler.

#### **2.2.1.5. Çoklu Lojistik Regresyon**

Bağımlı değişken üzerinde her zaman bir bağımsız değişken etkili olmayıp birden fazla değişkenin etkisinin olduğu durumlar da olabilmektedir. Bu durumlarda çoklu lojistik regresyon analizinden yararlanılmaktadır. Bu analiz yapılmadan önce

bağımsız değişkenlerin tek tek incelenerek ilgisiz değişkenlerin analize sokulmaması tavsiye edilmiştir. İncelemeler genel olarak iki yolla yapılmakta olup bunlardan ilkinde, bağımlı ve bağımsız değişkenler arasında Mann Whitney U, ki-kare, iki ortalama farkın önemlilik testi gibi testlerden faydalanmaktadır. İkincisinde ise bağımsız değişkenleri birer birer tek değişkenli lojistik regresyon analizi yaparak  $p < 0,25$  çıkan bağımsız değişkenin çoklu lojistik regresyona alınması önerilmektedir (13,14).

Çok değişkenli regresyon modeli Eşitlik (19)'da ki gibi gösterilir.

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (19)$$

#### **2.2.1.5.1. Etki Karışımı ve Etkileşim**

Etki karıştırıcı değişkenin varlığından söz edebilmek için sonuç değişken ile etkisi araştırılan bağımsız değişken arasında ilişkili olan bir ortak değişkenin olması gerekir. Böyle bir ortak değişkenin var olması durumunda sonuç değişken ile açıklayıcı değişken arasındaki ilişkinin karıştığı bu karışımı ortadan kaldırmak içinde etki karıştırıcı değişkenin modele eklenmesi gerektiği söylenebilir. Kurulan regresyon modelinde bağımlı değişkenin bir bağımsız değişkenle ilişkisi başka bağımsız değişkenin düzeyine göre değişiyorsa burada etkileşimden bahsedilir (13,14).

#### **2.2.1.5.2. Çoklu Bağlantı Sorunu**

Bağımsız değişkenler arasında bir ya da birden fazla bağlantının olması çoklu bağlantı (multicollinearity) problemini ortaya çıkarır. Mükemmel çoklu bağlantı olduğunda modeldeki katsayılar elde edilemezler. Bu durum genelde değişkenler arasında yüksek korelasyondan ( $R > 0.93$ ) kaynaklanır. Değişkenler arasındaki düşük düzeydeki bağlantı problemi genel olarak sorun olmaz. Ama yüksek düzeydeki bir ilişki, aslında etkisinin önemli olan bir değişkenin modele olan etkisini istatistiksel açıdan önemsiz bulunmasına sebep olur.

Çoklu bağlantı sorunu ile karşılaşıldığı zaman, değişkenlerden birisini modele dâhil etmek, değişkenlerden yeni bir değişken oluşturarak onu dâhil etmek ya da kısıtlanmış lojistik regresyon analizinden faydalanmak gibi yöntemlere başvurulabilir (13,37–39).

### 2.2.1.5.3. Sıfır Frekans ve Tam Ayırma Sorunu

Sıfır frekans ve tam ayırma sorunu çoklu bağlantı sorununda olduğu gibi regresyon katsayıları ve güven aralıklarının aşırı büyük kestirilmesine sebep olabilirler. Niteliksel bağımsız değişkenlerin birkaç kategorilerinde gözlemlediğimiz frekansların sıfır olması veya bağımlı değişkenin niteliksel bir değişken tarafından tam ayrılması durumunda sıfır frekans ve tam ayırma sorunu ortaya çıkabilir. Beklenen frekanslar sıfır olmayıp, 1'den büyük olmalıdır. 5'ten küçük olan ise toplam gözlemin %20'sini geçmemelidir (40).

Bağımlı değişken, bağımsız değişken tarafından çok iyi bir şekilde tahmin ediliyorsa tam ayrışım durumu vardır (41). Özellikle gözlem sayısının az ve değişken sayısının çok olduğu zaman ortaya çıkan sıfır frekans ve tam ayırma sorununu yaptığımız çalışmadaki gözlem sayıları arttırılarak kategorilerde yeterli frekansların olması sağlanıp ya da kategorik değişkenin bazı kategorileri birleştirilerek giderilmeye çalışılır (13).

### 2.2.1.6. Model Seçim Stratejisi

Bir veri setinden en iyi yaklaşımı oluşturabilmek için bir model seçim süreci gerekir. Verilerin en iyi açıklandığı ve en az karmaşıklığa sahip nihai bir modele ulaşılması gerekir. Bulunan bütün modeller en iyi model bulunana kadar elenerek gider. Son kalan model çıkarım için kullanılır.

Model seçim kriterleri, en iyi tahmin modelini, ilişkilendirme modelini ve sınıflandırma modelini seçmeye yarar. Model seçiminin birçok yöntemi vardır. Bu yöntemlerin başında akaike bilgi kriteri (AIC), bayesci bilgi kriteri (BIC) ve doğru sınıflama oranları gelir. Araştırmacının kendi amacına uygun olacak şekilde modeli seçmesi gerekse de AIC ve BIC düşük doğru sınıflama oranları yüksek olan modeli seçmek faydalıdır (42).

İlk olarak kullanılacak model seçim yöntemine karar verilmelidir. Lojistik regresyon analizinde iki temel yöntem vardır. Bunlar: standart (enter) ve adımsal (stepwise) yöntemlerdir. Adımsal yöntemler, ileriye doğru (forward) ve geriye doğru (backward) yöntemler olmak üzere ikiye ayrılır. Kimi araştırmacılar adımsal yöntemlerin verideki rastlantısal değişimlerden etkilendiğini, modeli aynı örnekler ile

tekrar ettiklerinde az da olsa aynı sonuçlar elde edilmesinden dolayı standart yöntemin sadece teori test ederken kullanılmasının uygun olduğunu iddia etmektedir (43). Field (2009)'a göre “bir bağımsız değişkenin diğerinden daha önemli olduğu düşünülüyorsa standart yöntem kullanılabilir”.

#### **2.2.1.6.1. Tam Regresyon Yöntemi**

Tüm bağımsız değişkenler aynı anda modele dâhil edilirler. Bağımsız değişkenler arasında üstünlük olduğu düşünülmez. Değişken sayısının artmasıyla tanımlayıcı modellerin sayısı da artacak ve bu değişkenlerin anlamlılıklarının belirlenme süresi uzayıp zaman kaybına sebep olacaktır. Belirtilen bu sebeplere dayanarak diğer yöntemlerin daha kullanışlı hale geldi söylenebilmektedir (44).

#### **2.2.1.6.2. İleriye Yönelik Seçim Yöntemi**

Modelden değişkenlerin çıkarılması ya da eklenmesi için değişkenlerin önemini kontrol eden bir yöntemdir (45).

Alpar (2013) bu yöntemi “bağımsız değişkenlerin olmadığı sadece sabitin olduğu, model ile işleme başlanır. Modele eklendiğinde log olabilirlik üzerinde en fazla değişime neden olan bağımsız değişken belirlenir ve bu değişken modele alınır. Eklenen değişkenin modele katkısı önemsiz bulunana kadar modele bağımsız değişkenler eklenmeye devam edilir” şeklinde ifade etmektedir (13).

#### **2.2.1.6.3. Geriye Dönük Eleme Yöntemi**

Modelin tüm bağımsız değişkenler kullanılarak birlikte kurulması (tam model) ve sırasıyla anlam ifade etmeyen bağımsız değişkenlerin modelden çıkarılması işlemidir. Adım adım modelden değişkenler atılarak işleme en son sapmada en çok değişikliğe neden olan değişken bulunduğu anda yöntem tamamlanmış olur (46).

#### **2.2.1.6.4. Adım-Adım Regresyon**

Her adımda modelin anlamlılığını test ederek değişken seçer. Wald ile  $\beta$  katsayılarının anlamlılıklarına göre seçilen değişkenler bir adımda modele ilave

edilirken sonraki adımlarda çıkarılabilirler. Anket çalışması gibi model kurma çalışmalarında iyi sonuç vermektedir (47).

#### 2.2.1.7. Lojistik Regresyonda Açıklayıcılık Katsayıları ( $R^2$ )

Doğrusal regresyonda, modelin açıklama gücünü gösteren  $R^2$ , bağımlı değişkendeki farklılaşmanın bağımsız değişkenler tarafından ne kadarının açıklanabildiğini gösteren bir ölçüdür. Doğrusal regresyonda  $R^2$  Eşitlik (20)'deki gibi hesaplanır. Modelin açıklama gücü  $R^2$  1'e yaklaştıkça artar (29).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (20)$$

Burada,

$n$ : gözlem sayısı,

$y_i$ : bağımlı değişken,

$\bar{y}$ : bağımlı değişken  $y$ 'nin ortalaması,

$\hat{y}_i$ : model yardımıyla kestirilen değerdir.

Lojistik regresyon analizi için birkaç  $R^2$  istatistiği bulunmaktadır. Sözde (pseudo) McFaden  $R^2$ , Cox-Snell ve Nagelkerke  $R^2$  istatistikleri en sık kullanılanlarıdır (13,29).

- **McFaden  $R^2$**

$$R_{McF}^2 = 1 - \frac{\ln \hat{L}(M_\beta)}{\ln \hat{L}(M_\alpha)} \quad (21)$$

$M_\alpha$ : sabit terimli modelin log-olabilirliği

$M_\beta$ : bağımsız değişkenlerin olduğu modelin log-olabilirliği

0,20 ile 0,40 arasındaki  $R^2$  değerleri LRA' da çok yüksek kabul edilir.

- **Cox-Snell  $R^2$**

$$R_{CS}^2 = 1 - \left\{ \frac{\hat{L}(M_\alpha)}{\hat{L}(M_\beta)} \right\}^{\frac{2}{n}} \quad (22)$$

En küçük 0 değerini alırken en büyük değeri 1 olmaz bu yüzden yorumlanması zor bir yöntemdir.

- **Nagelkerke  $R^2$**

$$R_{NAG}^2 = \frac{R_{CS}^2}{R_{MAX}^2} = \frac{1 - \left\{ \frac{\hat{L}(M_\alpha)}{\hat{L}(M_\beta)} \right\}^{2/n}}{1 - \hat{L}(M_\alpha)^{2/n}} \quad (23)$$

Nagelkerke 'nin önerdiği düzeltme ile Cox-Snell  $R^2$ 'nin en büyük değeri 1 olmaktadır. Bunun için Cox-Snell  $R^2$  değeri maksimum  $R^2$  değerine bölünmelidir. Böylelikle Nagelkerke  $R^2$  1'e eşit olur (48).

### 2.2.1.8. Model Uyum İyiliğinin Değerlendirilmesi

Uygun yöntemlerle değişkenler modele dâhil edildikten sonra yapılması gereken oluşturulan bu modelin uyum iyiliğinin değerlendirilmesidir. Anlamalı değişkenler kadar model uyumu da önem ifade eder. Uyum iyiliğinde bakılması gereken oluşturulan modelin verilere ne kadar iyi uyduğudur (49). Diğer bir deyişle modelin bağımlı değişkeni tanımlamada ne derecede iyi olduğunu inceler.

Model uyum iyiliğinin, modelde aykırı değerlerin olması, modelin tanımlanmasının yanlış yapılması, yanlış ölçek türünün kullanılması gibi durumlara istinaden test edilmesi gerekmektedir (50). Lojistik regresyonda oluşturulan modelde normallik varsayımı gerekli şart olmadığı için parametrik testler yerine nonparametrik testler kullanılır. Model uyum iyiliğinin değerlendirilmesinde Pearson ki-kare ve sapma testi, Hosmer-Lemeshow testi ve sınıflandırma tablosu en çok tercih edilenlerdir.

#### 2.2.1.8.1. Pearson ki-kare ve Sapma (Deviance) İstatistikleri

Modelin genel uyumunun sınanmasında kullanılan Pearson ve sapma artıkları gözlenen ve tahmin edilen değerler arasındaki farkların ölçüleridir. Pearson istatistiği ve sapma istatistiği, k bağımsız değişken sayısı (k-1) serbestlik derecesi ile ki-kare dağılımı gösterir. Pearson ve sapma istatistikleri (kritik) tablo değerinden büyük ve  $p < 0,05$  ise modelin uyumlu olmadığı söylenir (22,50).

Pearson ki-kare istatistiği Eşitlik (24), Sapma istatistiği Eşitlik (25)'te verilmiştir.

$$\chi^2 = \sum_{i=1}^J r_i^2 \quad (24)$$

$$D = \sum_{i=1}^J d_i^2 \quad (25)$$

#### 2.2.1.8.2. Hosmer ve Lemeshow Testi

Gözlenen ve beklenen olasılıklar arasındaki farkla birlikte genel uyumun değerlendirilmesi için hazırlanmış olan Hosmer-Lemeshow (H-L) testi 1980-1982 yıllarında David Hosmer ve Stanley Lemeshow tarafından tasarlanmıştır (51). Testin sonuçlarının anlamlı çıkmaması kurulan modelle verinin uyumlu olduğunu gösterir (52).

$$\hat{C} = \sum \frac{(G-B)^2}{B} \quad (26)$$

$G$ : Gözlenen Değer

$B$ : Beklenen Değer

Eşitlik (26) k-2 serbestlik derecesinde ki-kare dağılımına yaklaşır. Bu yöntem için en az 400 gözlem olması önerilmektedir. Ki-kare değerlerinin büyük olması modelin uyumunun iyi olmadığını gösterir (13).

Analiz çıktılarında Hosmer-Lemeshow değeri yer almazsa (SPSS), modelin ki-kare değerinin orijinal -2log likelihood (-2LL) değerine bölünmesiyle elde edilir.

#### 2.2.1.8.3. Sınıflandırma Tablosu

Lojistik regresyon analizinde uyum iyiliği değerlendirilmesinde kullanılan yöntemlerden birisi de sınıflandırma tablolarıdır. Bağımlı değişken 0 ve 1 ile kodlanarak değişkenin gözlenen ile beklenen değerlerin çaprazlanması ile oluşur. Çoğu zaman kesim noktası olarak 0,5 değeri alınır. Y eksenini 0,5 noktasının üzerindeki değişkenler 1'e, 0,5 noktasının altında kesinler 0 değerine atanır. Duyarlılık, seçicilik ve doğruluk değerleri ne kadar yüksek ise model uyumu o kadar iyidir (13,50).

**Tablo 2.3** Örnek Sınıflandırma Tablosu (Alpar, 2013).

| Gözlenen    | Kestirilen  |             | Doğru Sınıflama Yüzdesi         |              |
|-------------|-------------|-------------|---------------------------------|--------------|
|             | Pozitif (1) | Negatif (0) |                                 |              |
| Pozitif (1) | 56          | 7           | $56/(56+7)*100=88,9$            | (Duyarlılık) |
| Negatif (0) | 22          | 65          | $65/(65+22)*100=74,7$           | (Seçicilik)  |
| Toplam      |             |             | $(56+65)/(56+7+22+65)*100=80,7$ | (Doğruluk)   |

#### 2.2.1.9. Model Yeterliliği İle İlgili Ölçüler

Elde ettiğimiz sonuç modeli kullanmadan önce modelin yeterliliğinin saptanması gerekir. Bu saptamayı yapabilmek için kullanılabilecek ölçüler vardır. Bunların arasında en sık kullanılanları artıkların incelenmesi, açıklayıcılık katsayılarının belirlenmesi, çoklu bağlantı sorununun tespiti, artık grafiklerinin çizimidir.

##### 2.2.1.9.1. Artıkların İncelenmesi

Gerçek bağımlı değişken değeri  $y_i$  ve tahmin edilen bağımlı değişken değeri  $\hat{y}_i$  arasında ki uyumun bir göstergesi olan artıklar lojistik regresyonda normal dağılım göstermemesi ve değişen varyanslılık sorunu (heteroscedasticity) olması sebebiyle lojistik regresyonda artık incelenmesi doğrusal regresyona göre daha komplikedir (53,54). Pearson ve sapma artıkları sık kullanılsa da bunlara ek olarak lojit ve student türü artıklarda kullanılmaktadır.

**Tablo 2.4.** Artık Değerler ve Uzaklıklar (Field, 2009).

| Değer ve Artıklar         | Adı                                                                | Yorum                                                                                                                                                                            |
|---------------------------|--------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PRE_1<br>PGR_1<br>COO_1   | Tahmin edilen değer<br>Tahmin edilen grup<br>Cook Uzaklığı         | 1'den küçük olmalıdır.                                                                                                                                                           |
| LEV_1                     | Uzaklık (Leverage)                                                 | 0 (etki yok) ve 1 (tam etki) arasında değer alır. Beklenen uzaklık $(k+1)/N$ , k yordayıcı sayısı ve N örneklem büyüklüğü.                                                       |
| SRE_1<br>ZRE_1            | Standartlaştırılmış Artıklar                                       | Yalnızca %5'i +1,96'nın ve %1'i +2,58 dışında olabilir. Bu şekilde 3'ün üzerinde durum olduğunda endişe edilmeli, 3'e yakın durum olduğunda da, bu durumlar kontrol edilmelidir. |
| DEV_1<br>DFB0_1<br>DFB1_1 | Sapma (Deviance)<br>Sabit için DFBeta<br>İlk yordayıcı için DFBeta | 1'den küçük olmalıdır.                                                                                                                                                           |

**Ham Artıklar**

Bağımsız değişken(ler) yardımıyla kestirilen olasılık  $\hat{\pi}_i = P(y_i = 1 | x_i)$  olmak üzere  $e_i$  Eşitlik (25) ile tanımlanır.

$$e_i = y_i - \hat{\pi}_i \quad (27)$$

Artıklar binom dağılım özelliği gösterirken varyansları  $\hat{\pi}_i$  değerine bağlı olarak değişir (13).

**Pearson Artıkları**

Ham artık değerini kendi standart sapmasına bölerek Pearson artık değerine ulaşılır. Pearson değeri -2 ile +2 arasında olmayan gözlemler aşırı gözlemlerdir (55).

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}} \quad (28)$$

### Standartlaştırılmış Pearson ve Sapma Artıkları

Homojen varyansa sahip olmayan artıklar standart hatasına bölünerek standartlaştırılır. Standartlaştırılmış artıkların ortalaması 0, standart hataları ise 1 olur (56). Pearson artıklarının hesaplanması Eşitlik (29)'da, sapma artıklarının hesaplanması Eşitlik (30)'da verilmiştir.

$$r_{std_i} = \frac{r_i}{\sqrt{1-h_{ii}}} \quad (29)$$

$$d_{std_i} = \frac{d_i}{\sqrt{1-h_{ii}}} \quad (30)$$

### Lojit Artıklar

$$l_i = \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)} \quad (31)$$

### Student Türü Artıklar

Regresyon analizi sonuçlarına bağlı olarak hesaplanır. Örneklem büyüklüğü arttıkça student türü artıklar yaklaşık normal dağılmaktadır. Genellikle -2 ile +2 arasında değer alması beklenir (13,55).

#### 2.2.1.9.2. Etkili Gözlemlerin Belirlenmesi

##### Gözlem Uzaklığı (Leverage $h_{ii}$ )

Uzak gözlemler  $h_{ii}$  değerinin  $(p+1)/N$  değerinden büyük olduğu gözlemler için kullanılır. Gözlem sayısı 30'un altında olan istatistikler için  $3p/n$ , daha büyük gözleme sahip istatistikler için  $2p/n$  değerlerinin kesim noktası olarak kabul edilmesi önerilmiştir.

##### Cook Uzaklığı

Cook' un (1977) önerdiği etkili gözlemlerin değerlendirmesinde kullanılan cook uzaklığı her bir gözlemin tahmin değerlerine etkisi ölçülerek bulunur. Cook uzaklığı Eşitlik (32)'deki gibi bulunur.

$$D_j = \frac{z_j^2 h_j}{1-h_j} \quad (32)$$

Bir gözlemi sildiğimiz zaman diğer gözlemlerin ne kadar etkilendiğini ölçmek için cook uzaklığından yararlanırız.  $D_j > 1$  eşik değerini aşan gözlemler etkili gözlemlerdir (57). Genel bir kural olarak 0,5'ten büyük olan Cook uzaklık değerleri gözden geçirilmelidir.

### DFBETA

DFBETA istatistiği Belsley vd. (1980) tarafından önerilen, gözlem silme tekniğine dayalı olan tanı istatistiklerindendir (58). DFBETAS teşhis ölçüsü birinci gözlem silindiğinde regresyon katsayılarında ve standart sapmadaki değişimi gösteren bir ölçüdür (57). Diğer gözlemlerden anlamlı bir şekilde değişiklik gösteren DFBETA değerine sahip gözlemler etkili gözlem olarak isimlendirilebilirler (13).

#### 2.2.1.10. Katsayıların Yorumlanması

Lojistik regresyonda eğim katsayısı ( $\beta_1$ ), bağımsız değişkendeki bir birimlik değişimin lojitte ne kadarlık değişime sebep olduğunu gösterir.

Lojistik regresyon modelinde,

$$g(x + 1) = \beta_0 + \beta_1(x + 1), \quad (33)$$

$$g(x) = \beta_0 + \beta_1 x \quad (34)$$

$$g(x + 1) - g(x) = \beta_1 \text{ kadarlık değişime sebep olur.} \quad (35)$$

olmak üzere;

$$g(x + 1) - g(x) = \beta_1 \text{ kadarlık değişime sebep olur.} \quad (35)$$

Katsayıların yorumlanması için bağımsız değişkenin türünü bilmemiz gerekir.

Bağımsız değişken niteliksel ya da sürekli sayısal bir değişken olabilir.

#### Niteliksel Bir Bağımsız Değişkenin Yorumlanması

Bağımsız değişkenler niteliksel ve iki sonuçlu olduğu zaman katsayıların yorumlanması odds oranı ile yapılır ve lojistik regresyon modelinin aldığı değerler Tablo 2.5'teki gibi hesaplanır.

**Tablo 2.5.** Bağımsız Değişken İki Kategorili Olduğunda Lojistik Regresyon Modelinin Olasılık Değerleri

| Bağımsız Değişken | Bağımlı Değişken                                               |                                              |
|-------------------|----------------------------------------------------------------|----------------------------------------------|
|                   | x = 1                                                          | x = 0                                        |
| $\gamma=1$        | $\pi(1) = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}}$ | $\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$ |
| $\gamma=0$        | $1 - \pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$                 | $1 - \pi(0) = \frac{1}{1+e^{\beta_0}}$       |
| Toplam            | 1                                                              | 1                                            |

$x = 1$  için odds;

$$\frac{\pi(1)}{1-\pi(1)} \quad (36)$$

$x = 0$  için;

$$\frac{\pi(0)}{1-\pi(0)} \quad (37)$$

olarak hesaplanır.  $x = 1$  için odds,  $x = 0$  için olan odds'a bölünürse

$$OR = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \quad (38)$$

olur ve bu orana odds oranı denir.

$$g(1) = \ln\left(\frac{\pi(1)}{1-\pi(1)}\right) \quad g(0) = \ln\left(\frac{\pi(0)}{1-\pi(0)}\right) \text{ olmak üzere;} \quad (39),(40)$$

$$\ln(OR) = \ln\left(\frac{\pi(1)}{1-\pi(1)}\right) - \ln\left(\frac{\pi(0)}{1-\pi(0)}\right) = g(1) - g(0) \text{ olarak bulunur.} \quad (41)$$

$OR$ 'ye Tablo 2.5'teki değerler yazılırsa;

$$OR = \frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} / \frac{1}{1+e^{\beta_0+\beta_1}}}{\frac{e^{\beta_0}}{1+e^{\beta_0}} / \frac{1}{1+e^{\beta_0}}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (42)$$

buradan  $\ln(OR) = \beta_1$  olarak bulunur. Yani iki kategorili niteliksel bağımsız değişkenin lojistik regresyon için üstünlük oranı  $e^{\beta_1}$  dir. Logit farkta  $\beta_1$  'e eşit olacaktır (59).

Üstünlük oranı ( $OR$ ) yaygın kullanılan bir ölçümdür.  $x=1$  ' in  $x=0$  'a göre olma olasılığının ne kadar çok veya ne kadar az olduğunu kestirmemize yarar (60).  $OR$  ve lojistik regresyon parametresi sonuçları yorumlamada bakacağımız önemli çıktılardır. Sonuçların daha kolay yorumlanıp anlaşılabilmesi için odds oranı  $\beta_1$  katsayısının e üssü (exponansiyeli) alınarak  $e^{\beta_1}$  hesaplanır (13).

Örneğin aile öyküsünde diyabeti olan kişilerin diyabet hastalığına yakalanma riski araştırılacak olursa, yapılan lojistik regresyon analizi sonucunda aile öyküsünde diyabeti olma değişkeninin katsayısı  $\beta_1=0,642$  olarak hesaplandığını kabul edelim. Buna göre “aile öyküsünde diyabeti olan kişilerin diyabet hastalığına yakalanma riski 1,9 kat( $e^{0,642}=1,9$ ) daha fazladır” şeklinde yorum yapılır.

### **Sürekli Sayısal Bağımsız Değişkenin Yorumlanması**

Sürekli sayısal bağımsız değişkenler için odds oranı yorumlanırken, lojitin sürekli değişkenle arasında doğrusal bir ilişki olmalıdır (14). Bu doğrusallık sağlanmış olduğunda:

$$g(x + 1) - g(x) = [\beta_0 + \beta_1(x + 1)] - [\beta_0 + \beta_1x] = \beta_1 \quad (43)$$

$\beta_1$  eğim katsayısı, bağımsız değişkendeki ( $x$ ) 1 birimlik değişimin lojitte meydana getirdiği farkı gösteriyor. Lojistik regresyon modelinde sürekli bağımsız değişken olduğu zaman parametrelerin yorumlanması, değişkenin modele nasıl alındığına ve değişkenin birimine (cm, m, kg, vb.) göre değişiklik gösterir (61).

Örneğin yaş değişkeninde 1 birimlik artışlar çok anlam ifade etmeyeceğinden 10 birimlik artışlardan bahsedilir. Öte yandan sayısal bağımsız değişken 0 ve 1 arasında değer alan oranlar olduğunda da 1 birimlik artış çok büyük olup 0,10 birimlik artışlar için yorum yapmak daha mantıklı olacaktır (13).

$x$ 'deki  $c$ 'lik artış için lojistik fonksiyonunda meydana gelen fark;

$$g(x + c) - g(x) = c\beta_1 \text{ olur.} \quad (44)$$

Odds oranı da bu farkın e üssü alınarak  $e^{c\beta_1}$  bulunur. Odds oranı bulunduktan sonra güven aralığı hesaplanabilir. *OR*, *β1* katsayısı için güven aralığı aşağıdaki gibi hesaplanır (62).

$$e^{(c\hat{\beta}_1 \pm z_{1-\alpha/2} cS(\hat{\beta}_1))} \quad (45)$$

Seçilen c değerine bağlı olarak güven aralıkları ve nokta kestirimi değişmektedir. Örneğin yaş değişkeni ele alarak bu durumu anlatalım. Lojistik regresyon katsayısının 0,06 olduğu, standart hatanın 0,025 olduğunu varsayarak lojittteki 1 birimlik değişimin riskte 1,06 kat ( $e^{0,06}=1,06$ ) arttığı hesaplanabilir. Odds oranı ile %95 güven aralığında alt sınır Eşitlik(46)'da, üst sınır Eşitlik (47)'deki gibi hesaplanır.

$$e^{0,06-1,96*0,025} = 1,01 \quad (46)$$

$$e^{0,06+1,96*0,025} = 1,11 \quad (47)$$

Yaştaki 10 birimlik artış ise riski 1,49 kat ( $e^{10*0,06} = 1,49$ ) artmasına sebep olacaktır. 10 birimlik artışın standart hatası 0,25 ( $10*0,025$ ) olur. Güven aralığının alt sınırı Eşitlik (48)'de, üst sınır Eşitlik (49)'daki gibi hesaplanır.

$$e^{0,6-1,96*0,25} = 1,11 \quad (48)$$

$$e^{0,6+1,96*0,25} = 2,97 \quad (49)$$

### 2.2.2. Çok Kategorili (Multinomial) Lojistik Regresyon Analizi

Bağımlı değişkenin herhangi bir sıra olmadan üç ve daha fazla kategoriden oluştuğu zaman çok kategorili lojistik regresyon analizinden faydalanılır. Bağımlı değişken 4 kategoriye sahipse “4-1” adet lojistik regresyon modeline sahip olması gerekmektedir (63). Çünkü bir kategori referans kategorisidir. Diğer kategoriler bu referans kategorisine göre incelenir (13). Örneğin referans kategori 0 ise karşılaştırmalar 0-1, 0-2, 0-3 olmak üzere 3 lojit model oluşur (64). Bu modeller aşağıdaki eşitlikler gibidir.

$$\ln \left[ \frac{P(y=1|x_1)}{P(y=0|x_1)} \right] = \beta_1 + \beta_{11}x_1 \quad (50)$$

$$\ln \left[ \frac{P(y=1|x_2)}{P(y=0|x_1)} \right] = \beta_2 + \beta_{21}x_1 \quad (51)$$

$$\ln \left[ \frac{P(y=1|x_3)}{P(y=0|x_1)} \right] = \beta_3 + \beta_{31}x_1 \quad (52)$$

İki kategorili lojistik regresyonda uygulanan modele bağımsız değişkenlerin alınma prosedürü, model parametrelerin kestirimi, değişkendeki önemliliklerinin değerlendirilmesi ve sonuçların yorumlanması çok kategorili lojistik regresyon içinde geçerlidir (13).

Çok kategorili lojistik regresyon modelinde lojit dönüşümde iki kategorili lojistik regresyondaki dönüşüm ile benzerlik gösterir. Kaç tane lojit model oluşursa o kadar odds oranı hesaplanır ve bu oranların doğal logaritmaları alınarak lojit modeller elde edilir ve model doğrusal hale getirilmiş olur (65).

Tüm bağımsız değişkenler modele alınabilir. Bazı istatistik analiz yapan programlarda çok kategorili lojistik regresyon için adımsal seçim yöntemi yoktur. Onun için referans kategori ile diğer kategoriler sırasıyla ele alınır.

Çok kategorili lojistik regresyonda bağımsız değişkenler arasında yüksek ilişki olmaması, tam ayrılma sorunu olmaması veya sıfır frekans sorununun olmaması için gözlem sayısının bağımsız değişken sayısının 10 katı olması önerilmiştir (13).

### 2.2.3. Sıralı (Ordinal) Lojistik Regresyon Analizi

Üç ve daha fazla kategorili sıralanabilir niteliksel durumda bağımlı değişken olduğu zaman sıralı lojistik regresyon analizi yapılır. Sıralanabilir olması demek örneğin bir hastalığın evreleri (1-2-3), düşük-orta-yüksek, iş performansı (olağanüstü-tatmin edici- yetersiz) vb. diye kodlayabildiğimiz durumlar bağımlı değişkenin kendi içinde sıralanabilirliğidir. Bağımsız değişken sıralanabilir olduğu halde sırayı gözetmeksizin çok kategorili regresyon yapıldığı zaman bilgi kaybına ve tahmin modelinin yanlış yorumlanmasına neden olunmaktadır (14,66).

Sıralı regresyon modelinin temel özellikleri;

1. Sonuç değişkeni gizli bir değişkenin etkisi altındadır. Bu gizli değişken sıralı ve gruplanmış kategorik değişken şeklindedir. Bu kategoriler arası gerçek uzaklıklar tam olarak hesaplanamaz (67).
2. Bağlantı fonksiyonu kullanılır. Bu fonksiyonun normal dağılımlı ve sabit varyanslı olmasına gerek yoktur (68).
3. Bağımsız değişken ile bağımlı değişken arasındaki ilişki kategorilerden bağımsızdır. Çünkü regresyon katsayısı değeri de kategorilerden bağımsızdır. Başka bir deyişle model şunu varsayar: link fonksiyonundaki ilgili regresyon katsayıları her kesme noktası için eşittir (69).

Orantısal odds lojistik modeli aşağıdaki gibi tanımlanır;

$$g_j(x) = \ln \left[ \frac{p(Y \leq j|x)}{p(Y > j|x)} \right] \quad j=0, 1, 2, \dots, J \text{ için} \quad (53)$$

Ardışık kategori lojistik regresyon modeli Eşitlik (54)'te ki gibidir.

$$g_j(x) = \ln \left[ \frac{\theta_j(x)}{\theta_{j-1}(x)} \right] = \alpha_j + x'\beta \quad j=0, 1, 2, \dots, J \text{ için} \quad (54)$$

Bu modelin varsayımı olarak, bağımsız değişkenin etkisini yorumlamak için alınan kesim noktaları farklılık gösterse bile odds oranlarının aynı olmasıdır. Burada modeller arasındaki paralellik sınanır ve bu sıralı lojistik regresyon için en belirgin özelliştir.

Sıralı lojistik regresyon modelinde yer alan gizli değişken, gözlemlediğimiz kategorik bağımlı değişkeni etkileyen,  $(-\infty, +\infty)$  arasında değer alan gözlemlenemeyen gizli bir sonuç değişkeninden bahsetmek için kullanılır (25).

Bağlantı fonksiyonu bir olasılık fonksiyonudur. Tamamlayıcı loglog, negatif loglog, cauchit, logit ve probit olmak üzere 5 farklı bağlantı fonksiyonu vardır. En sık logit ve probit fonksiyonları kullanarak sıralı lojistik regresyon modeli elde edilir (66).

### 3. GEREÇ VE YÖNTEM

#### 3.1. Araştırmanın Modeli

Bu çalışma, retrospektif tarzda tasarlanmış olup açık erişime sahip iki farklı veri seti ile tamamlanmıştır. Birinci örnek veri setinde iki kategorili (binary) lojistik regresyon modeli için Cleveland Clinic verisi olarak bilinen ve araştırmacılar tarafında sıklıkla tercih edilen kalp krizi riski veri kümesi kullanılmıştır. Hungarian Institute of Cardiology (Budapeşte/Macaristan), University of Hospital (Zürih/İsviçre), University Hospital (Basel/İsviçre) ve V.A. Medical Center (Cleveland/ABD) sağlık kurumlarında çok merkezli olarak derlenen veri kümesi toplamda 76 değişkenden oluşmaktadır. Ancak değişkenlerin büyük bir bölümü veri kümesinden çıkarılmış ve 14 değişken ile daha sade ve kompakt bir veri kümesi haline dönüştürülmüştür.

İkinci örnek veri setinde ise çok kategorili (multinomial) lojistik regresyon modeli için sağlık sigortası verisi kullanılmıştır. Veri kümesi, ABD’de yaşayan kişilere ait sigorta poliçelerinde bulunan bilgilerden idari kayıt olarak derlenmiştir.

#### 3.2. Araştırmanın Yapıldığı Grup

Kalp krizi riski veri setinde yer alan hastalar göğüs ağrısı nedeniyle sağlık merkezlerine başvuran ve yaş aralığı 29 ile 77 yıl arasında olan kadın ve erkek hastalardır. Yaş medyanı 55 yıl olarak hesaplanmıştır. Verilerin toplanması çok merkezli olarak dört farklı hastanede gerçekleştirilmiştir.

Sağlık sigortası harcaması veri setinde ise ABD’de yaşayan kadın ve erkek müşterilerin yapmış oldukları sağlık sigortası poliçelerinden elde edilen bilgiler ile derlenmiştir. Müşterilerin yaş aralığı 18 ile 64 yıl arasında değişmektedir. Yaş medyanı 39 yıl olarak hesaplanmıştır.

#### 3.3. Verilerin Toplanması

Çalışmada kullanılan veri kümeleri web ortamında açık erişime sahip ve genel kullanıma açık olan bir siteden alınmıştır. Veri kümeleri, sahipleri tarafından kullanıma izin verilmiş olup, aşağıdaki bağlantılar kullanılarak veri kümeleri ile ilgili ayrıntılı bilgi edinilebilir:

Birinci veri seti için: <https://www.kaggle.com/ronitf/heart-disease-uci>

İkinci veri seti için: <https://www.kaggle.com/mirichoi0218/insurance>

Veri kümelerine ilişkin değişkenler şöyledir:

### **Kalp Krizi Riski Veri Seti**

- A) Yaş
- B) Cinsiyet
- C) Göğüs ağrısı tipi;
  - Değer 1: Tipik anjina pektoris
  - Değer 2: Atipik anjina pektoris
  - Değer 3: Anjin dışı ağrı
  - Değer 4: Aseptomatik
- D) Dinlenme pozisyonu kan basıncı
- E) Serum kolesterol (mg/dL)
- F) Açlık kan şekeri (AKŞ) >120 ve <120 mg/dL
- G) Dinlenme pozisyonu Elektrokardiyografi (EKG) bulguları
  - Değer 0: Normal
  - Değer 1: ST-T dalgası anormalliği var (T dalgası inversiyonları / veya > 0.05 mV ST yükselmesi veya depresyonu)
  - Değer 2: Olası veya kesin sol ventrikül hipertrofisini (Estes kriterlerine göre)
- H) Ulaşılan maksimum kalp hızı
- I) Egzersize bağlı anjina(EBA) (1 = evet; 0 = hayır)
- J) Dinlenmeye göre egzersizin neden olduğu ST çökmesi
- K) Pik egzersizin eğimi ST segmenti
  - Değer 1: Yukarı eğim
  - Değer 2: Düz
  - Değer 3: Aşağı eğimli
- L) Floroscopy ile renklendirilen ana damar sayısı
- M) 1:normal, 2:sabit kusur, 3:tersine çevrilebilir kusur
- N) 0:Daha düşük kalp krizi riski  
1:Daha yüksek kalp krizi riski

## **Sağlık Sigortası Harcama Veri Seti**

- A) Yaş
- B) Cinsiyet
- C) Vücut Kitle İndeksi(VKİ)
- D) Sigortadan faydalanan çocuk sayısı
- E) Sigara kullanımı
- F) Sağlık sigorta harcaması (USD)
- G) İkamet bölgesi
  - Güneydoğu (1)
  - Güneybatı (2)
  - Kuzeydoğu (3)
  - Kuzeybatı (4)

### **3.4. Verilerin Analizi**

Çalışmanın istatistiksel analizleri SPSS 20.0 (IBM InCorp, Chicago, IL, USA) programı ile gerçekleştirilmiştir. Tanımlayıcı istatistiksel ölçüler kategorik veriler için sıklık ve yüzde oranı; sayısal veriler için ortalama $\pm$ SS(Standart Sapma) ve dağılımın çarpık olduğu durumlarda (medyan; min-maks) şeklinde sunulmuştur. İki bağımsız gruba göre karşılaştırmalarda Bağımsız Grup Student t-testi veya Mann-Whitney U testi kullanılmıştır. Çoklu gruplara göre yapılan karşılaştırmalarda Tek Yönlü Varyans Analizi (ANOVA) veya Kruskal-Wallis testi kullanılmıştır.

Anlamlı bulunan karşılaştırma sonuçları için ikili karşılaştırmalarda Tukey HSD post-hoc testi tercih edilmiştir. Kategorik veriler arasındaki ilişkilerin saptanması amacıyla ki-kare analizi uygulanmıştır. Kalp krizi riski veri seti için düşük ve yüksek risk sınıfları üzerinde anlamlı etkisi olan değişkenlerin belirlenmesi amacıyla iki kategorili (binary) lojistik regresyon modeli; sağlık sigorta harcaması veri kümesinde ülkenin dört farklı bölgesinde yaşayan sigorta müşterilerinin özelliklerinin etkisini belirlemek amacıyla çok kategorili (multinomial) lojistik regresyon modelleri kurulmuştur.

Çalışmada önem seviyesi (significance level)  $\alpha=0,05$  olarak alınmıştır. Sonuçların daha kolay izlenmesi amacıyla tablolar ve görselleştirilmesi amacıyla ilgili grafikler oluşturulmuştur.



## 4. BULGULAR

### İki Kategorili (Binary) Lojistik Regresyon Modeli

Çalışmaya göğüs ağrısı şikâyeti ile hastane veya sağlık merkezi acil bölümü veya kardiyoloji bölümüne başvuran toplam 303 hasta dâhil edilmiştir. Hastaların yarısından fazlası erkektir (%68,32; n=207) (Şekil 4.1).

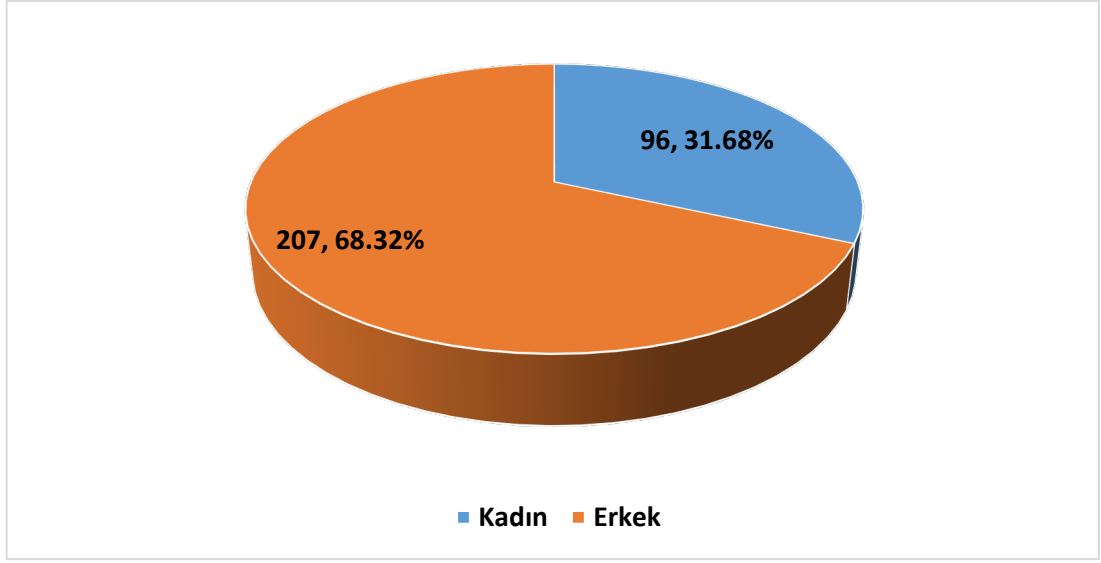
Hastaların yarıya yakınında (%47,2; n=143) önemli semptom görülmezken %16,5 oranında tipik anjina pektoris, %28,7 oranında atipik anjina pektoris ve geri kalanında (%7,6) anjina dışı ağrı bulunmaktadır (Şekil 4.2).

Hastaların %14,9'unda açlık kan şekeri (AKŞ) 120 mg/dL üzerinde bulunmuştur. Hastaların yarıya yakınında (%48,5) dinlenme pozisyonunda EKG bulguları normal iken büyük oranda (%50,2) ST-T dalga anormalliği saptanmıştır. Düşük oranda (%1,3) sol ventrikül hipertrofisi gözlenmiştir.

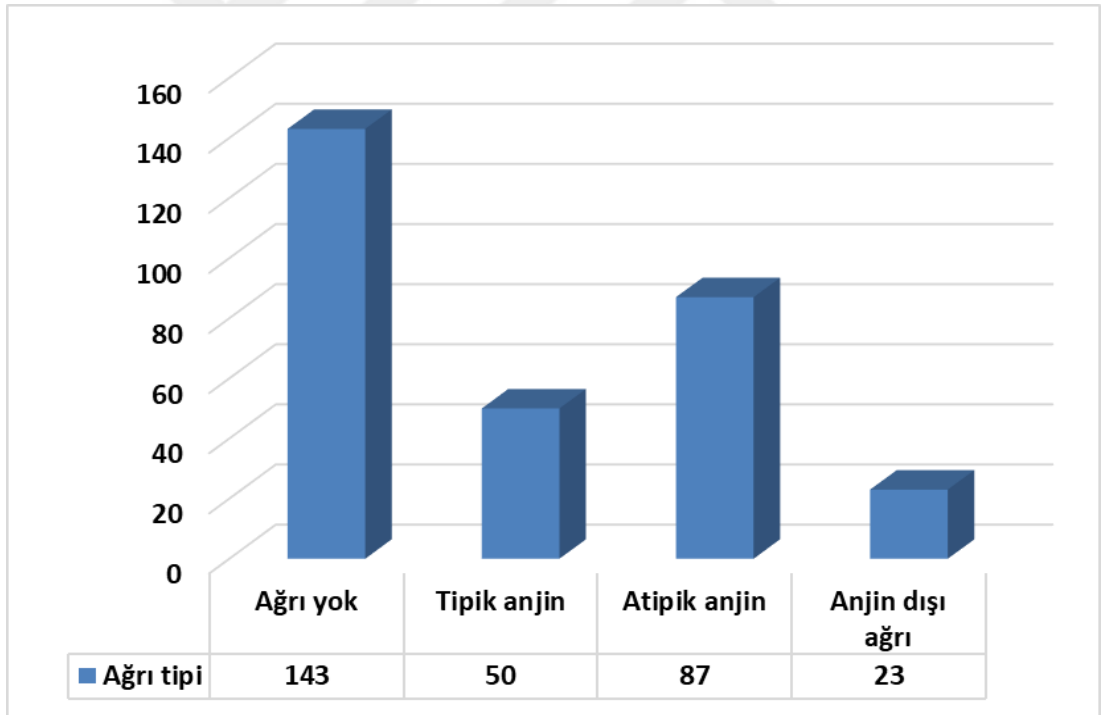
Hastaların yaklaşık üçte birinde (%32,7) egzersiz ile birlikte gelişen anjina pektoris (AP) bulunmaktadır. Egzersizin pik anında ST segmentinin eğimi %46,2 oranında yukarı doğru, %46,9 oranında ise düz devam etmiştir. Düşük oranda aşağı doğru eğim bulunmaktadır.

Hastaların %45,5'inde düşük myokard infarktüsü (MI) riski, yarısından fazlasında ise (%54,5) yüksek MI riski bulunmaktadır (Şekil 4.3). Hastaların yaş ortalaması 54,36±9,08 yıl (55; 29-77 yıl) olarak hesaplanmıştır. Dinlenme dönemi kan basıncı değeri 131,62±17,53 mmHg (94-200) olurken maksimum ulaşılan nabız sayısı 149,64±22,90 (71-202) atım/dk olmuştur.

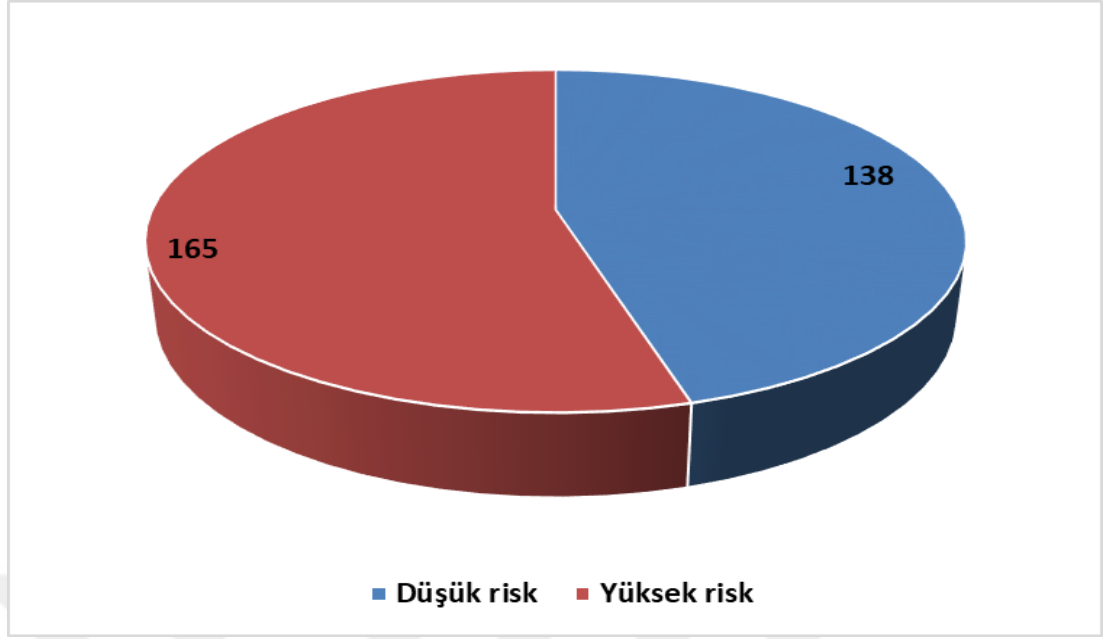
Hastaların total kolesterol değeri 246,26±51,83 mg/dL olup 126 ile 564 mg/dL arasında değişmektedir. ST segment çökmesi ölçümleri ortalaması 1,16±0,80 (0-6,2) bulunmuştur. Floroscopy ile görüntülenen ana damar sayısı 0 ile 4 arasında değişmektedir.



Şekil 4.1. Çalışmaya Katılan Hastaların Cinsiyet Oranları



Şekil 4.2. Çalışmaya Katılan Hastaların Başvuru Şikâyetleri



**Şekil 4.3.** Çalışmaya Alınan Hastaların Kalp Krizi Risk Sınıfları

Hastaların yaşı, dinlenme anı kan basıncı, maksimum nabız sayısı, ST segmenti çökme değeri ve görüntülenen damar sayısı, düşük ve yüksek MI risk sınıfları arasında anlamlı farklılık göstermiştir. Ayrıca ki-kare analiz sonuçlarına göre erkek hastaların oranı yüksek MI risk grubunda anlamlı düzeyde yüksek olup OR değeri 1,669 olarak hesaplanmıştır.

AKŞ oranları risk grupları arasında anlamlı bulunmamıştır, ancak AKŞ için risk değeri OR=1,077 olarak hesaplanmıştır. Dinlenme anı EKG bulgularında ST-T dalga anormalliği yüksek risk grubunda bulunan hastalarda daha yüksek oranda gözlenmiştir. Egzersiz anında anjina görülme oranı düşük risk grubundaki hastalarda daha yüksek oranda izlenmiştir.

Lojistik regresyon modeli için risk sınıfları bağımlı değişken olarak belirlenmiş ve düşük risk grubu referans kategori olarak belirlenmiştir.

### **Modelin Uygulanması**

Veri setinde verilen bilgilerden hareketle hastaların düşük MI ve yüksek MI risk sınıflarının tahmin edilebileceği düşünülmüştür. Bu değişkende 0: düşük risk ve 1: yüksek risk olarak kodlanmıştır. Bu durum, iki kategorili lojistik regresyon modelinin kurulması için uygundur.

Lojistik regresyon temelde iki amaç için kullanılır: belirli bir kategoriye tahmin etmek için veya bağımsız değişkenlerin ilgili kategoriler üzerindeki etki düzeyini hesaplamak için. Bu çalışmada, lojistik regresyon modellemesi hastalara ait bağımsız değişkenlerin sınıf bilgisi olan düşük ve yüksek MI riski üzerindeki etkilerini belirlemek üzere yapılacaktır.

Lojistik regresyon modeli oluşturulurken iki temel model seçimi yapılır. Birinci seçim tam lojistik modelidir ve “Enter” yöntemi olarak adlandırılır. İkinci seçim ise tüm regresyon modellerinde genel olan ileriye dönük seçim, geriye dönük elimine etmek ve aşamalı yöntemlerdir.

Lojistik regresyon modellerinde bu yöntemlerin farklı çeşitleri de bulunmaktadır. Aşamalı seçim yöntemi uygulama için kullandığımız SPSS programında bulunmayıp ileriye dönük ve geriye dönük yöntemlerin koşullu, Wald ve olabilirlik türleri bulunmaktadır. Kalp krizi riskine yönelik veri seti için “Enter” yöntemi ile “İleriye dönük: Olabilirlik” yöntemlerine göre modellerin oluşturulmasına karar verilmiştir. “Enter” yöntemi ile oluşturulan tam lojistik model için 0: düşük MI riski kategorisi referans kategori olarak belirlenmiş ve diğer tüm bağımsız değişkenler modele dâhil edilmiştir.

Regresyon modellerinde genel olarak modelin anlamlılığı, bağımsız değişkenlerin anlamlılığı, modelin uyumu, model seçimi ve modelin tahmin etmedeki başarısı konuları incelenir. Bu testlerin hepsini başarılı bir şekilde geçen modellerin kullanılabileceğine karar verilir. Gerek etki belirleme gerekse tahmin (predictive) modeli oluşturma amacıyla olsun başarılı modeller bir soruna çözüm olması için kullanılır ve elde edilen bilgiler oldukça kıymetlidir. Ayrıca lojistik regresyon “Enter” yöntemi ile oluşturulan modele ilişkin tablolar aşağıda sırasıyla ve detaylı bir biçimde sunulmuştur.

**Tablo 4.1.** Yapılan Analize İlişkin Gözlem Sayısı Bilgileri

| <b>Süreç Özeti</b>         |                      |          |          |
|----------------------------|----------------------|----------|----------|
| <b>Ağırlıksız Durumlar</b> |                      | <b>n</b> | <b>%</b> |
| Seçilmiş Durumlar          | Analize Dahil Edilen | 303      | 100,0    |
|                            | Eksik Gözlem         | 0        | 0,0      |
|                            | Toplam               | 303      | 100,0    |
| Seçilmemiş Durumlar        |                      | 0        | 0,0      |
| Toplam                     |                      | 303      | 100,0    |

İlk tablo yapılan analiz ile ilgili modele dâhil edilen gözlem değerleri, dâhil edilmeyen gözlemler ve eksik gözlemlerin sayısı ve yüzde oranları bilgisini sunmaktadır. Veri kümesinde bulunan 303 hastaya ait kayıtların tamamının modele katıldığı, herhangi bir eksik gözlem olmadığı anlaşılmaktadır.

Regresyon modellerinde genel olarak katsayıların tahmin edilmesinde ( $X'X$ ) kareler toplamı ve iç çarpımlar (SSCP) matrisi oluşturulur ( $X$ : bağımsız değişkenlerden oluşan veri matrisi). Eğer gözlem değerlerinde eksik gözlem varsa bu durumda  $X'X$  matrisinde çarpım yapılacak değer olmayacaktır. Bu durumda matrisin rank değeri azalır. Bu durum, katsayıların yanlış tahmin edilmesine yol açabilir.

**Tablo 4.2.** Bağımlı Değişken Kodlama Sembolü Bilgisi

| <b>Bağımlı Değişken Kodlanması</b> |                        |
|------------------------------------|------------------------|
| <b>Orijinal Kategori</b>           | <b>Kodlama Sembolü</b> |
| Düşük Risk                         | 0                      |
| Yüksek Risk                        | 1                      |

Bağımlı değişkenin kodlama durumunu gösteren küçük bir tablo bilgisi oluşturulur. Bağımlı değişken olarak belirlediğiniz değişkene ait kodlanmış olan kategorilerin neler olduğu ve kodlamanın ne şekilde yapıldığı bilgisi görülür.

Genel olarak iki kategorili lojistik regresyon modellerinde 0 ve 1 kodlarının olması gerekir ancak bazı durumlarda veri girişi 1 ve 2 şeklinde veya metin olarak “evet/hayır”, “var/yok”, “hasta/sağlam” şeklinde giriş yapılmaktadır. Eğer veri setinde

kodlama yapılmazsa program kodlarda “1” varsa onu model kategorisi olarak alıp diğerini referans kategorisi olarak belirlemektedir.

Bu tabloda buna ait bilgi sunulmaktadır. Tablo 4.2’de çalışmaya ait bağımlı değişken bilgisi ve kategorileri sorunsuz bir biçimde tanımlanmıştır.

**Tablo 4.3.** Modele Alınan Kategorik Değişkenlerin Sıklık ve Kodlama Bilgileri

| <b>Kategorik Değişkenlerin Kodlanması</b> |                            |               |                          |            |            |
|-------------------------------------------|----------------------------|---------------|--------------------------|------------|------------|
|                                           |                            | <b>Sıklık</b> | <b>Parametre Kodları</b> |            |            |
|                                           |                            |               | <b>(1)</b>               | <b>(2)</b> | <b>(3)</b> |
| Göğüs Ağrısı                              | Aseptomatik                | 143           | 1,000                    | 0,000      | 0,000      |
|                                           | Tipik Anjin                | 50            | 0,000                    | 1,000      | 0,000      |
|                                           | Atipik Anjin               | 87            | 0,000                    | 0,000      | 1,000      |
|                                           | Anjin Dışı Ağrı            | 23            | 0,000                    | 0,000      | 0,000      |
| ST Eğimi                                  | Aşağı Eğim                 | 21            | 1,000                    | 0,000      |            |
|                                           | Yukarı Eğim                | 140           | 0,000                    | 1,000      |            |
|                                           | Düz                        | 142           | 0,000                    | 0,000      |            |
| Dinlenme EKG                              | Normal                     | 147           | 1,000                    | 0,000      |            |
|                                           | ST-T Dalga Anormalliği     | 152           | 0,000                    | 1,000      |            |
|                                           | Sol Ventrikül Hipertrofisi | 4             | 0,000                    | 0,000      |            |
| AKŞ                                       | Düşük <120                 | 258           | 1,000                    |            |            |
|                                           | Yüksek >120                | 45            | 0,000                    |            |            |
| EBA                                       | Yok                        | 204           | 1,000                    |            |            |
|                                           | Var                        | 99            | 0,000                    |            |            |
| Cinsiyet                                  | Kadın                      | 96            | 1,000                    |            |            |
|                                           | Erkek                      | 207           | 0,000                    |            |            |

Lojistik regresyon modellerinin çoklu doğrusal regresyon modellerinde olduğu gibi katı varsayımları yoktur. Bunlardan birisi de modele kategorik değişkenin eklenebilmesidir. Doğrusal regresyon modelinde bu durum yalnızca kukla (dummy)

değişken eklenmesi ile mümkün olurken, lojit dönüşümü ile lojistik modellere kategorik değişken eklenmesi mümkündür.

Çalışmamızda modele eklenen kategorik bağımsız değişkenler Tablo 4.3'te gösterilmiştir. Kategorik değişkenlerin ilk kodlanan kategorileri referans kategori olarak belirlenmiştir.

**Tablo 4.4.** Gözlenen ve Tahmin Edilen Sınıflandırma (Çapraz Geçerlilik) Tablosu

| <b>Sınıflandırma Tablosu</b> |                   |             |                      |                    |                                    |
|------------------------------|-------------------|-------------|----------------------|--------------------|------------------------------------|
| <b>Gözlenen</b>              |                   |             | <b>Tahmin Edilen</b> |                    |                                    |
|                              |                   |             | <b>Kalp Krizi</b>    |                    | <b>Düzeltilmiş<br/>Yüzde Oranı</b> |
|                              |                   |             | <b>Düşük Risk</b>    | <b>Yüksek Risk</b> |                                    |
| Aşama 0                      | Kalp Krizi        | Düşük Risk  | 0                    | 138                | 0,0                                |
|                              |                   | Yüksek Risk | 0                    | 165                | 100,0                              |
|                              | Genel yüzde oranı |             |                      |                    | 54,5                               |

Lojistik regresyonda modelin başarısı ile ilgili bilgiyi veren tablo sınıflandırma tablosudur. Tahmin modellerinde bu tabloya “hata matrisi” denilir. Gözlenen kategori oranlarını ve model oluşturulduktan sonra tahmin edilen kategorilerin oranlarını çapraz tablo şeklinde verir. Genel yüzde oranı bilgisi modelin tahminlemedeki başarı oranını gösterir.

Çalışmamızda başlangıç modeli için “yüksek MI riski” kategorisi tahmin başarısı %54,5 bulunmuştur (Tablo 4.4). Lojistik regresyonda modelin oluşturulması iteratif şekilde ilerler. Başlangıç aşaması olarak modele yalnızca sabit terim eklenir, diğer bağımsız değişkenler eklenmez. Başlangıç aşamasında modelin sabit terimi için beta katsayısı anlamlı bulunmamıştır ( $p=0,121$ ).

Lojistik regresyon modellerinde bağımsız değişkenlere ait katsayıların anlamlılık testi Wald skorunun anlamlılık sınaması ile gerçekleştirilir (Tablo 4.5). Wald testi, modele ilişkin katsayı eğimlerinin sıfıra eşit olup olmadığını test eden bir hipotez testidir.

Lojistik regresyon modelinde katsayıların anlamlılığı ile ilgili skor testi de kullanılır. Skor testi başlangıç aşamasında hesaplanır ve modele hangi değişkenlerin anlamlı katkı yaptığı ile ilgili bilgi verir. Veri seti için başlangıç aşaması değişkenlere

ait skor testi sonuçları verilmiştir (Tablo 4.6). Buna göre, kolesterol, AKŞ ve ST yukarı eğim kategorilerinin modele anlamlı katkı yapmayacağı anlaşılmıştır ( $p>0,05$ ).

**Tablo 4.5.** Başlangıç Aşaması Model Bilgileri

| Modeldeki Değişkenler |       |       |       |       |    |       |        |
|-----------------------|-------|-------|-------|-------|----|-------|--------|
|                       |       | B     | SH    | Wald  | sd | p     | Exp(B) |
| Aşama 0               | Sabit | 0,179 | 0,115 | 2,400 | 1  | 0,121 | 1,196  |

**Tablo 4.6.** Modele Eklenmeyen Değişkenlerin Skor Bilgileri

| Modele Eklenmeyen Değişkenler |                  |                                           |         |    |          |
|-------------------------------|------------------|-------------------------------------------|---------|----|----------|
|                               |                  |                                           | Skor    | sd | p        |
| Aşama 0                       | Değişkenler      | Yaş                                       | 15,399  | 1  | $<0,001$ |
|                               |                  | Cinsiyet (Erkek)                          | 23,914  | 1  | $<0,001$ |
|                               |                  | Göğüs Ağrısı                              | 81,686  | 3  | $<0,001$ |
|                               |                  | Göğüs Ağrısı (tipik)                      | 80,680  | 1  | $<0,001$ |
|                               |                  | Göğüs Ağrısı (atipik)                     | 18,318  | 1  | $<0,001$ |
|                               |                  | Göğüs Ağrısı (anjin dışı)                 | 30,399  | 1  | $<0,001$ |
|                               |                  | Dinlenme Kan Basıncı                      | 6,365   | 1  | 0,012    |
|                               |                  | Kolesterol                                | 2,202   | 1  | 0,138    |
|                               |                  | AKŞ ( $>120$ mg/dL)                       | 0,238   | 1  | 0,625    |
|                               |                  | Dinlenme EKG                              | 10,023  | 2  | 0,007    |
|                               |                  | Dinlenme EKG (ST-T dalga anormallliği)    | 7,735   | 1  | 0,005    |
|                               |                  | Dinlenme EKG (Sol ventrikül hipertrofisi) | 9,314   | 1  | 0,002    |
|                               |                  | Maksimum Nabız                            | 53,893  | 1  | $<0,001$ |
|                               |                  | Egzersiz Anjin                            | 57,799  | 1  | $<0,001$ |
|                               |                  | ST Çökmesi                                | 56,206  | 1  | $<0,001$ |
|                               |                  | ST Eğim                                   | 47,507  | 2  | $<0,001$ |
|                               |                  | ST Eğim (yukarı)                          | 1,224   | 1  | 0,269    |
|                               |                  | ST Eğim (düz)                             | 39,718  | 1  | $<0,001$ |
|                               |                  | Damar Sayısı                              | 46,495  | 1  | $<0,001$ |
|                               |                  | Kusur                                     | 35,862  | 1  | $<0,001$ |
|                               | Genel İstatistik |                                           | 161,493 | 17 | $<0,001$ |

Omnibus testi, oluşturulan modelin genel anlamlılığı ile ilgili sınamanın yapıldığı ve hipotez sonucuna karar verildiği bir yöntemdir. Test istatistiği ki-kare

dağılımına uyar ve modele eklenen tüm katsayıların serbestlik derecesine göre ki-kare tablo değerine göre hipotez testi gerçekleştirilir. Sıfır hipotezinin reddedilmesi, modelin anlamlılığı anlamına gelmektedir.

Doymuş (tam) model oluşturulduğunda Omnibus testindeki aşama, blok ve model için hesaplanan olabilirlik ki-kare değerleri eşit bulunur. Ancak ileriye veya geriye dönük modeller kullanıldığında her bir iterasyon için farklı ki-kare değeri hesaplanır. “Enter” seçimi ile oluşturulan modelimizde LR Ki-kare=211,53 ve  $p<0,001$  bulunduğundan modelin anlamlı olduğuna karar verilir (Tablo 4.7).

**Tablo 4.7.** Modele İlişkin Omnibus Testi

| <b>Model Katsayılarının Omnibus Testi</b> |       |                |           |                 |
|-------------------------------------------|-------|----------------|-----------|-----------------|
|                                           |       | <b>Ki-kare</b> | <b>sd</b> | <b><i>p</i></b> |
| Aşama 1                                   | Aşama | 211,535        | 17        | $<0,001$        |
|                                           | Blok  | 211,535        | 17        | $<0,001$        |
|                                           | Model | 211,535        | 17        | $<0,001$        |

Lojistik regresyon modelinde uyum iyiliği testlerinde birisi de açıklayıcılık katsayıları olan  $R^2$  değerleridir. Buradaki açıklayıcılık değerleri doğrusal regresyondaki gibi yorumlanmaz. Ancak yine de katsayıların modele yapmış oldukları katkı hakkında fikir vermektedir.  $R^2$  değerinin yüksek olması modelin daha iyi uyuma sahip olduğunu gösterir.

Öncelikle -2LL değerine bakılır. Küçülen değerler modelin uyumunda iyileşme olduğu anlamına gelmektedir. Ancak doymuş modelde bir tane -2LL değeri olacağından bir karşılaştırma yapmak mümkün değildir. -2LL=206,103 bulunmuş olup modelin uyumu ile ilgili daha farklı bilgilere ihtiyaç vardır. Bu anlamda, Cox & Snell  $R^2$  ve Nagelkerke  $R^2$  değeri incelenir. Genel olarak Nagelkerke katsayısı Cox & Snell katsayısından daha yüksek bulunur.

Lojistik regresyon modellerinde genel olarak 1’e yakın  $R^2$  değerleri elde edilmez. Çalışmaya ait olan  $R^2$  katsayıları 0,502 ve 0,672 olarak hesaplanmış olup, oldukça iyi uyum düzeyinden söz edilebilir (Tablo 4.8).

**Tablo 4.8. Model Açıklayıcılık Değerleri**

| <b>Model Özeti</b> |                          |                                      |                                 |
|--------------------|--------------------------|--------------------------------------|---------------------------------|
| <b>Aşama</b>       | <b>-2 Log likelihood</b> | <b>Cox &amp; Snell R<sup>2</sup></b> | <b>Nagelkerke R<sup>2</sup></b> |
| 1                  | 206,103                  | 0,502                                | 0,672                           |

Modelin anlamlılığı ile ilgili Hosmer & Lemeshow testi uygulanır. Bu testin hipotezleri “Model anlamlıdır” ve “Model anlamlı değildir” şeklinde kurulur. Sıfır hipotezinin reddedilmediği durumlarda modelin anlamlı olduğuna karar verilir. Bu test özellikle risk tahmini için kullanılır. Verilerin modele ne kadar uygunluğunu gösterir.

Serbestlik derecesi 8 olan ki-kare istatistiği 8,866 olarak hesaplanmış ve  $p=0,354$  olarak hesaplandığından anlamlı bulunmamıştır. Bu durumda sıfır hipotezi reddedilemez ve modelin anlamlı olduğuna karar verilir (Tablo 4.9). Doymuş (tam) modelde ilk aşamadaki sınıflandırma tablosuna göre düşük risk için %77,5 ve yüksek risk tahmin başarısı %92,1 ve modelin genel tahmin başarısı %85,5 hesaplanmıştır (Tablo 4.10).

**Tablo 4.9. Hosmer & Lemeshow Model Uyum İyiliği**

| <b>Hosmer and Lemeshow Testi</b> |                |           |          |
|----------------------------------|----------------|-----------|----------|
| <b>Aşama</b>                     | <b>Ki-kare</b> | <b>sd</b> | <b>p</b> |
| 1                                | 8,866          | 8         | 0,354    |

**Tablo 4.10. Sınıflandırma Tablosu**

| <b>Sınıflandırma Tablosu</b> |                   |             |                      |                    |                                        |
|------------------------------|-------------------|-------------|----------------------|--------------------|----------------------------------------|
| <b>Gözlenen</b>              |                   |             | <b>Tahmin Edilen</b> |                    |                                        |
|                              |                   |             | <b>Kalp Krizi</b>    |                    | <b>Düzeltilmiş<br/>Yüzde<br/>Oranı</b> |
|                              |                   |             | <b>Düşük Risk</b>    | <b>Yüksek Risk</b> |                                        |
| Aşama 1                      | Kalp Krizi        | Düşük Risk  | 107                  | 31                 | 77,5                                   |
|                              |                   | Yüksek Risk | 13                   | 152                | 92,1                                   |
|                              | Genel Yüzde Oranı |             |                      |                    | 85,5                                   |

“Enter” yöntemi ile oluşturulan tam (doymuş) model için değişkenlerin katsayıları ve anlamlılık değerleri hesaplanmıştır. Wald istatistiği sınaması anlamlılık bulunan değişkenler tespit edilmiştir.

Lojistik regresyon modeli oluşturulduğunda modelin katsayıları, anlamlılık sınamaları, risk değerleri ve güven aralıkları ile bir bütün olarak hesaplanır. Hesaplanan katsayının anlamlılığı iki yolla incelenir. İlki Wald istatistiğinin anlamlılık değeridir. İkincisi ise OR değerine ait güven aralığının “1” değerini içerip içermediğidir. Beta katsayısının üstel değeri (exponential) OR değerini yani risk ya da koruyucu faktör olarak değerlendirme değerini verir.

Hasta yaşının yüksek kalp krizi üzerinde bir risk faktörü olmadığı görülmektedir ( $p=0,932$ ). Cinsiyetin erkek olması kalp krizi için yüksek risk faktörü olarak bulunmuştur. Beta katsayısı 1,83 ve  $\text{Exp}(1,83)=6,234$  ( $p<0,001$ ) olarak bulunmuştur.

OR değeri referans kategoriye göre bağımlı kategori üzerindeki riskin etkisini gösterir.  $\text{OR}=6,234$  için “Erkek cinsiyet, düşük kalp krizi riskine göre yüksek kalp krizi riskini 6,23 kat artırmaktadır” anlamına gelmektedir. OR için hesaplanan %95 güven aralığı ise OR değerinin hangi aralıkta olursa güvenilir olacağını göstermektedir.

Cinsiyet değişkeni için %95 güven aralığı 2,402 ile 16,179 bulunmuştur. Yani erkek cinsiyet kalp krizi riskini 2,4 kat ile 16,1 kat arasında artıran bir risk faktörüdür. Beta katsayısı pozitif bulunan değişkenler için artırıcı risk faktörü, negatif olanlar için azaltıcı risk (koruyucu) faktörü denilir.

Göğüs ağrısı değişkeni genel olarak anlamlı etkiye sahip olup tipik, atipik ve anjin dışı göğüs ağrıları içerisinde yalnızca tipik anjin anlamlı etkiye sahip bulunmuştur ( $p=0,001$ ).  $\text{Beta}=-2,202$  ve  $\text{Exp}(\text{Beta})=\text{OR}=0,111$  olduğundan tipik anjinin yüksek riski azaltıcı bir etkisi bulunmuştur.

Negatif beta katsayıları için OR değeri 1’in altında bulunur. Bu durumda, risk faktörü olmadığına karar verilir ancak kaç kat azaltıcı etkisi olduğu hesaplanan OR değerinden kolaylıkla anlaşılmaz. Bu nedenle, bulunan OR değerinin çarpmaya göre tersi alınır ve “1/OR” değeri ile etki değeri belirlenir. Tipik anjin için  $\text{OR}=0,111$  ve

güven aralığı 1 değerini içermemektedir. Buna göre,  $1/0,111=9,009$  değeri hesaplanır ve tipik anjin değişkeninin yüksek kalp krizi riskini 9,009 kat azalttığı kararı verilir.

Dinlenme periyodu kan basıncı, AKŞ, serum kolesterol düzeyi, dinlenme periyodu EKG bulguları, maksimum nabız sayısı ve yukarı doğru olan ST eğim değişkenlerinin modele anlamlı katkısının olmadığı görülmüştür.

Bunun yanı sıra, egzersiz periyodunda anjin varlığı, düz seyreden ST eğimi, floroskopi ile görüntülenen damar sayısının daha az olması ve tersine çevrilebilir ST defekti anlamlı etkiye sahip bulunmuştur. Bu değişkenlere ait olan risk ve koruyucu faktörler benzer şekilde hesaplanır (Tablo 4.11).

**Tablo 4.11. Doymuş Model Değişkenleri ve Anlamlılık Değerleri**

| <b>Modeldeki Değişkenler</b>  |        |       |        |    |        |              |                           |        |
|-------------------------------|--------|-------|--------|----|--------|--------------|---------------------------|--------|
|                               | Beta   | SH    | Wald   | sd | p      | Exp(Beta)=OR | OR için %95 Güven Aralığı |        |
|                               |        |       |        |    |        |              | Alt                       | Üst    |
| Yaş                           | -0,002 | 0,023 | 0,007  | 1  | 0,932  | 0,998        | 0,953                     | 1,045  |
| Cinsiyet (Erkek)              | 1,830  | 0,487 | 14,147 | 1  | <0,001 | 6,234        | 2,402                     | 16,179 |
| Göğüs Ağrısı                  |        |       | 23,084 | 3  | <0,001 |              |                           |        |
| Göğüs Ağrısı (tipik)          | -2,202 | 0,645 | 11,669 | 1  | 0,001  | 0,111        | 0,031                     | 0,391  |
| Göğüs Ağrısı (atipik)         | -1,103 | 0,757 | 2,124  | 1  | 0,145  | 0,332        | 0,075                     | 1,462  |
| Göğüs Ağrısı (anjin dışı)     | -0,144 | 0,642 | 0,050  | 1  | 0,823  | 0,866        | 0,246                     | 3,051  |
| Dinlenme KB                   | -0,018 | 0,010 | 2,945  | 1  | 0,086  | 0,982        | 0,962                     | 1,003  |
| Kolesterol                    | -0,004 | 0,004 | 1,229  | 1  | 0,268  | 0,996        | 0,988                     | 1,003  |
| AKŞ                           | 0,016  | 0,549 | 0,001  | 1  | 0,977  | 1,016        | 0,346                     | 2,979  |
| Dinlenme EKG                  |        |       | 1,753  | 2  | 0,416  |              |                           |        |
| Dinlenme EKG (ST anormalliği) | 0,301  | 2,174 | 0,019  | 1  | 0,890  | 1,351        | 0,019                     | 95,794 |

**Tablo 4.12.** Doymuş Model Değişkenleri ve Anlamlılık Değerleri (devam)

|                                           |        |       |        |   |        |        |       |         |
|-------------------------------------------|--------|-------|--------|---|--------|--------|-------|---------|
| Dinlenme EKG (sol ventrikül hipertrofisi) | 0,777  | 2,172 | 0,128  | 1 | 0,720  | 2,176  | 0,031 | 153,568 |
| Maks. Nabız                               | 0,018  | 0,011 | 2,922  | 1 | 0,087  | 1,019  | 0,997 | 1,040   |
| Egzersiz Anjina                           | 0,849  | 0,421 | 4,075  | 1 | 0,044  | 2,338  | 1,025 | 5,335   |
| ST Çökmesi                                | -0,504 | 0,220 | 5,235  | 1 | 0,022  | 0,604  | 0,392 | 0,930   |
| ST Eğim                                   |        |       | 5,748  | 2 | 0,056  |        |       |         |
| Eğim (yukarı)                             | -0,295 | 0,916 | 0,104  | 1 | 0,747  | 0,744  | 0,124 | 4,484   |
| Eğim (düz)                                | -1,041 | 0,450 | 5,359  | 1 | 0,021  | 0,353  | 0,146 | 0,852   |
| Damar Sayısı                              | -0,804 | 0,198 | 16,453 | 1 | <0,001 | 0,447  | 0,303 | 0,660   |
| ST Kusur                                  |        |       | 84,104 | 2 | 0,001  |        |       |         |
| ST Kusur (Sabit)                          | 0,734  | 0,524 | 1,963  | 1 | 0,161  | 2,084  | 0,746 | 5,818   |
| ST Kusur (tersine çevrilebilir)           | 2,771  | 0,308 | 81,020 | 1 | <0,001 | 15,978 | 8,739 | 29,213  |
| Sabit                                     | 4,090  | 3,511 | 1,357  | 1 | 0,244  | 59,724 |       |         |

Lojistik regresyon yöntemi her ne kadar varsayım gerektirmese de model ile ilgili bazı sorunların incelenmesi gerekmektedir. Ayrıca, artıkların incelenmesi ve modelin yapısını bozan etkin veya etkili gözlemler varsa bunların belirlenmesi gereklidir. Öncelikle, katsayı tahminlerinin anlamlılığını ve modelin uyumunu olumsuz etkileyen çoklu bağlantı problemi araştırılmıştır.

Lojistik regresyon modellemesi içerisinde çoklu bağlantı problemini saptayacak direk bir test yöntemi bulunmamaktadır. Bu amaçla doğrusal regresyon model oluşturma bölümünden modele eklenen değişkenler ile ilgili çoklu bağlantı problemi test edilmiştir. Genellikle varyans değerini arttıran faktör ( Variance inflation factor, VIF) ve koşul indeksi değerleri incelenir. VIF değerleri arttıkça çoklu bağlantı (multicollinerity) probleminin düzeyinin arttığına karar verilir. Genel olarak 10 değerine yaklaşan veya geçen VIF değerleri çoklu bağlantı probleminin göstergesidir.

Çalışmamızda VIF değerleri genellikle 1,5 civarında hesaplanmış ve en yüksek VIF değerinin 1,706 olması nedeniyle modelde çoklu bağlantı problemi olmayabileceği düşünülmüştür. Ancak, diğer bir ölçüt olan koşul indeksi değerinin de incelenmesi ve koşul indeksi değerinin yaklaşık 30 değerinin üzerinde olmaması gerekmektedir. Ancak, indeks değeri 50,785 değerine kadar artış göstermiştir. Bu durumda, modelde çoklu bağlantı probleminin olduğuna karar verilir (Tablo 4.12). Bu durumda modele eklenen bağımsız değişkenler arasındaki korelasyonların incelenmesi gerekmektedir. Aralarında korelasyon değeri yüksek olan değişkenlerden bir tanesinin çıkarılması sorunu çözecektir. Modele daha az katkı yapan değişken tercih edilir. Ancak bu sorunu gidermenin farklı yöntemleri de bulunmaktadır. Bu yöntemler tez konumuzun dışında olduğundan yöntemlerin uygulanmasına yönelik analizlere yer verilmemiştir. Onun yerine, modeldeki değişken sayısını azaltacak model seçimi ile bu sorunun giderilmesine karar verilmiştir (Tablo 4.14 - 4.16).

**Tablo 4.13. Çoklu Bağlantı Problemi Göstergeleri**

|              | VIF   | Koşul İndeksi |
|--------------|-------|---------------|
| (Sabit)      |       | 1,000         |
| Yaş          | 1,443 | 3,035         |
| Cinsiyet     | 1,162 | 3,345         |
| Göğüs Ağrısı | 1,284 | 4,121         |
| Dinlenme KB  | 1,171 | 4,444         |
| Kolesterol   | 1,150 | 4,638         |
| AKŞ          | 1,087 | 5,768         |
| Dinlenme EKG | 1,061 | 5,859         |
| Maks. Nabız  | 1,614 | 11,008        |
| EBA          | 1,402 | 14,121        |
| ST Çökmesi   | 1,706 | 18,318        |
| ST Eğim      | 1,643 | 19,528        |
| Damar Sayısı | 1,203 | 28,746        |
| ST Kusur     | 1,147 | 50,785        |

Modelde olabilirlik oran yöntemine uygun ileriye dönük model seçimi (forward LR) tercih edilmiştir. Bu yöntem seçilerek oluşturulan modele ilişkin anlamlılık sonuçları Tablo 4.13'te gösterilmiştir. En uygun model 7. aşamada oluşturulmuştur.

Omnibus ki-kare deęerleri tm modellerde anlamlı bulunmuştur. Ařama, blok ve model ki-kare deęerleri anlamlı bulunmuştur ( $p < 0,05$ ). Yedinci ařamada elde edilen modelin  $R^2$  uyum ve aıklayıcılık deęerleri olduka yksek bulunmuştur ( $R^2_{\text{Cox-Snell}} = 0,485$  ve  $R^2_{\text{Nagelkerke}} = 0,648$ ). Bu deęerler ve sınıflandırma tablosuna gre modelin genel tahmin bařarısı, tam modele gre biraz daha dřk olmakla birlikte Hosmer & Lemeshow ki-kare deęeri ve -2LL deęeri daha yksek bulunmuştur.

**Tablo 4.14.** İleriye Dnk Modele İliřkin Omnibus Testi Sonuları

| Model Katsayıları İin Omnibus Testi |       |         |    |          |
|--------------------------------------|-------|---------|----|----------|
|                                      |       | Ki-kare | sd | <i>p</i> |
| Ařama 1                              | Ařama | 85,941  | 3  | $<0,001$ |
|                                      | Blok  | 85,941  | 3  | $<0,001$ |
|                                      | Model | 85,941  | 3  | $<0,001$ |
| Ařama 2                              | Ařama | 38,719  | 1  | $<0,001$ |
|                                      | Blok  | 124,660 | 4  | $<0,001$ |
|                                      | Model | 124,660 | 4  | $<0,001$ |
| Ařama 3                              | Ařama | 28,645  | 1  | $<0,001$ |
|                                      | Blok  | 153,305 | 5  | $<0,001$ |
|                                      | Model | 153,305 | 5  | $<0,001$ |
| Ařama 4                              | Ařama | 15,164  | 1  | $<0,001$ |
|                                      | Blok  | 168,469 | 6  | $<0,001$ |
|                                      | Model | 168,469 | 6  | $<0,001$ |
| Ařama 5                              | Ařama | 10,948  | 1  | $0,001$  |
|                                      | Blok  | 179,417 | 7  | $<0,001$ |
|                                      | Model | 179,417 | 7  | $<0,001$ |
| Ařama 6                              | Ařama | 15,205  | 2  | $<0,001$ |
|                                      | Blok  | 194,622 | 9  | $<0,001$ |
|                                      | Model | 194,622 | 9  | $<0,001$ |
| Ařama 7                              | Ařama | 6,372   | 1  | $0,012$  |
|                                      | Blok  | 200,994 | 10 | $<0,001$ |
|                                      | Model | 200,994 | 10 | $<0,001$ |

**Tablo 4.15. İleriye Dönük Model İçin Uyum İyiliği Ölçütleri**

| <b>Model Özeti</b> |                          |                                      |                                 |
|--------------------|--------------------------|--------------------------------------|---------------------------------|
| <b>Aşama</b>       | <b>-2 Log likelihood</b> | <b>Cox &amp; Snell R<sup>2</sup></b> | <b>Nagelkerke R<sup>2</sup></b> |
| 1                  | 331,697                  | 0,247                                | 0,330                           |
| 2                  | 292,978                  | 0,337                                | 0,451                           |
| 3                  | 264,333                  | 0,397                                | 0,531                           |
| 4                  | 249,169                  | 0,427                                | 0,570                           |
| 5                  | 238,221                  | 0,447                                | 0,597                           |
| 6                  | 223,016                  | 0,474                                | 0,634                           |
| 7                  | 216,644                  | 0,485                                | 0,648                           |

**Tablo 4.16. İleriye Dönük Model İçin Hosmer & Lemeshow Testi Sonuçları**

| <b>Hosmer and Lemeshow Testi</b> |                |           |          |
|----------------------------------|----------------|-----------|----------|
| <b>Aşama</b>                     | <b>Ki-kare</b> | <b>sd</b> | <b>p</b> |
| 1                                | 0,000          | 2         | 1,000    |
| 2                                | 5,047          | 8         | 0,753    |
| 3                                | 4,887          | 8         | 0,770    |
| 4                                | 10,834         | 8         | 0,211    |
| 5                                | 12,359         | 8         | 0,136    |
| 6                                | 13,003         | 8         | 0,112    |
| 7                                | 10,544         | 8         | 0,229    |

**Tablo 4.17. İleriye Dönük Model İçin Sınıflandırma Tablosu**

| Sınıflandırma Tablosu |                   |             |               |             |                            |
|-----------------------|-------------------|-------------|---------------|-------------|----------------------------|
| Gözlenen              |                   |             | Tahmin Edilen |             |                            |
|                       |                   |             | Kalp Krizi    |             | Düzeltilmiş<br>Yüzde Oranı |
|                       |                   |             | Düşük Risk    | Yüksek Risk |                            |
| Aşama 1               | Kalp Krizi        | Düşük Risk  | 104           | 34          | 75,4                       |
|                       |                   | Yüksek Risk | 39            | 126         | 76,4                       |
|                       | Genel Yüzde Oranı |             |               |             | 75,9                       |
| Aşama 2               | Kalp Krizi        | Düşük Risk  | 110           | 28          | 79,7                       |
|                       |                   | Yüksek Risk | 42            | 123         | 74,5                       |
|                       | Genel Yüzde Oranı |             |               |             | 76,9                       |
| Aşama 3               | Kalp Krizi        | Düşük Risk  | 106           | 32          | 76,8                       |
|                       |                   | Yüksek Risk | 26            | 139         | 84,2                       |
|                       | Genel Yüzde Oranı |             |               |             | 80,9                       |
| Aşama 4               | Kalp Krizi        | Düşük Risk  | 103           | 35          | 74,6                       |
|                       |                   | Yüksek Risk | 25            | 140         | 84,8                       |
|                       | Genel Yüzde Oranı |             |               |             | 80,2                       |
| Aşama 5               | Kalp Krizi        | Düşük Risk  | 108           | 30          | 78,3                       |
|                       |                   | Yüksek Risk | 22            | 143         | 86,7                       |
|                       | Genel Yüzde Oranı |             |               |             | 82,8                       |
| Aşama 6               | Kalp Krizi        | Düşük Risk  | 112           | 26          | 81,2                       |
|                       |                   | Yüksek Risk | 21            | 144         | 87,3                       |
|                       | Genel Yüzde Oranı |             |               |             | 84,5                       |
| Aşama 7               | Kalp Krizi        | Düşük Risk  | 108           | 30          | 78,3                       |
|                       |                   | Yüksek Risk | 16            | 149         | 90,3                       |
|                       | Genel Yüzde Oranı |             |               |             | 84,8                       |

Tüm aşamalara ait olan ve modele dâhil edilen değişkenler ait katsayı tahminleri aşağıda verilmiştir (Tablo 4.17).

İlk aşamada yalnızca göğüs ağrısı kategorilerinin modele alındığı ve tipik anjin kategorisinin modele anlamlı katkı yaptığı görülmüştür. Bir sonraki aşamada ST çökmesi değeri göğüs ağrısı ile birlikte modele alınmıştır. 3.aşamada floroscopy ile görüntülenen damar sayısı dâhil edilmiştir. Böylece en sonunda 7.aşamada en uygun model oluşturulmuştur.

Cinsiyet, göğüs ağrısı, ST çökme, ST eğim, damar sayısı, egzersiz periyodu anjin ve ST defekti değişkenleri anlamlı olarak modele dâhil edilmiştir. Erkek cinsiyet,

egzersiz dönemi anjin, sabit ve tersine çevrilebilir ST defekti değişkenleri pozitif etkiye sahip bulunmuştur. Katsayılar pozitif ve OR değerleri 1'den büyüktür. Tipik göğüs ağrısı, ST çökmesi, SAT düz eğim ve damar sayısı katsayıları negatif hesaplandığından OR değerleri 1'den küçük bulunmuştur. OR değerlerinin yorumlanması da yukarıda anlatıldığı üzere 1/OR hesaplamasıyla bulunabilir.

Modele anlamlı katkı yapan değişkenlerin modelden çıkarılması durumunda modelin anlamlılığındaki değişimin hangi düzeyde olacağını gösteren durumlar Tablo 4.18'de sunulmuştur. Tablodaki 7. aşamaya ait durumlar incelendiğinde değişkenlerin tamamı için -2LL'deki değişimlerin anlamlı olduğu görülmektedir. Bu durumda, anlamlı bulunan değişkenlerin tamamı modelde bulunmalıdır. Modele anlamlı katkı yapmayan diğer değişkenlere ilişkin anlamlılık değerleri de Tablo 4.19'da verilmiştir.

**Tablo 4.18. İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri**

| Modeldeki Değişkenler |                           |        |       |        |    |        |                      |                              |       |
|-----------------------|---------------------------|--------|-------|--------|----|--------|----------------------|------------------------------|-------|
|                       |                           | Beta   | SH    | Wald   | sd | p      | Exp(Bet<br>a)<br>=OR | OR için %95<br>Güven Aralığı |       |
|                       |                           |        |       |        |    |        |                      | Alt                          | Üst   |
| Aşama 1               | Göğüs Ağrısı              |        |       | 72,572 | 3  | <0,001 |                      |                              |       |
|                       | Göğüs Ağrısı (tipik)      | -1,808 | 0,491 | 13,578 | 1  | <0,001 | 0,164                | 0,063                        | 0,429 |
|                       | Göğüs Ağrısı (atipik)     | 0,690  | 0,584 | 1,395  | 1  | 0,237  | 1,993                | 0,635                        | 6,259 |
|                       | Göğüs Ağrısı (anjin dışı) | 0,517  | 0,525 | 0,971  | 1  | 0,324  | 1,677                | 0,600                        | 4,691 |
|                       | Sabit                     | 0,827  | 0,453 | 3,328  | 1  | 0,068  | 2,286                |                              |       |
| Aşama 2               | Göğüs Ağrısı              |        |       | 54,304 | 3  | <0,001 |                      |                              |       |
|                       | Göğüs Ağrısı (tipik)      | -2,224 | 0,556 | 15,985 | 1  | <0,001 | 0,108                | 0,036                        | 0,322 |
|                       | Göğüs Ağrısı (atipik)     | -0,336 | 0,654 | 0,264  | 1  | 0,607  | 0,714                | 0,198                        | 2,575 |
|                       | Göğüs Ağrısı (anjin dışı) | 0,036  | 0,582 | 0,004  | 1  | 0,951  | 1,036                | 0,331                        | 3,244 |

**Tablo 4.19.** İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri (devam)

|         |                                 |        |       |        |   |        |         |       |        |
|---------|---------------------------------|--------|-------|--------|---|--------|---------|-------|--------|
|         | ST Çökmesi                      | -0,893 | 0,162 | 30,376 | 1 | <0,001 | 0,410   | 0,298 | 0,563  |
|         | Sabit                           | 2,198  | 0,565 | 15,116 | 1 | <0,001 | 9,009   |       |        |
| Aşama 3 | Göğüs Ağrısı                    |        |       | 47,377 | 3 | <0,001 |         |       |        |
|         | Göğüs Ağrısı (tipik)            | -2,149 | 0,580 | 13,753 | 1 | <0,001 | 0,117   | 0,037 | 0,363  |
|         | Göğüs Ağrısı (atipik)           | -0,366 | 0,682 | 0,288  | 1 | 0,591  | 0,693   | 0,182 | 2,639  |
|         | Göğüs Ağrısı (anjin dışı)       | 0,196  | 0,608 | 0,104  | 1 | 0,747  | 1,217   | 0,370 | 4,006  |
|         | ST Çökmesi                      | -0,899 | 0,172 | 27,278 | 1 | <0,001 | 0,407   | 0,291 | 0,570  |
|         | Damar Sayısı                    | -0,809 | 0,161 | 25,133 | 1 | <0,001 | 0,445   | 0,325 | 0,611  |
|         | Sabit                           | 2,679  | 0,604 | 19,670 | 1 | <0,001 | 14,573  |       |        |
| Aşama 4 | Göğüs Ağrısı                    |        |       | 42,343 | 3 | <0,001 |         |       |        |
|         | Göğüs Ağrısı (tipik)            | -2,124 | 0,589 | 13,010 | 1 | <0,001 | 0,120   | 0,038 | 0,379  |
|         | Göğüs Ağrısı (atipik)           | -0,481 | 0,693 | 0,483  | 1 | 0,487  | 0,618   | 0,159 | 2,401  |
|         | Göğüs Ağrısı (anjin dışı)       | 0,199  | 0,617 | 0,104  | 1 | 0,747  | 1,221   | 0,364 | 4,093  |
|         | ST Çökmesi                      | -0,882 | 0,181 | 23,697 | 1 | <0,001 | 0,414   | 0,290 | 0,590  |
|         | Damar Sayısı                    | -0,814 | 0,171 | 22,718 | 1 | <0,001 | 0,443   | 0,317 | 0,619  |
|         | ST Kusur                        |        |       | 55,049 | 2 | <0,001 |         |       |        |
|         | ST Kusur (Sabit)                | 1,532  | 0,657 | 5,441  | 1 | 0,020  | 4,627   | 1,277 | 16,762 |
|         | ST Kusur (tersine çevrilebilir) | 2,991  | 0,404 | 54,753 | 1 | <0,001 | 19,913  | 9,016 | 43,979 |
|         | Sabit                           | 4,973  | 0,897 | 30,758 | 1 | <0,001 | 144,465 |       |        |
|         | Cinsiyet (Erkek)                | 1,240  | 0,390 | 10,131 | 1 | 0,001  | 3,456   | 1,610 | 7,418  |
|         | Göğüs Ağrısı                    |        |       | 42,196 | 3 | <0,001 |         |       |        |

**Tablo 4.20.** İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri (devam)

|         |                                 |        |       |        |   |        |        |       |        |
|---------|---------------------------------|--------|-------|--------|---|--------|--------|-------|--------|
| Aşama 5 | Göğüs Ağrısı (tipik)            | -2,283 | 0,593 | 14,799 | 1 | <0,001 | 0,102  | 0,032 | 0,326  |
|         | Göğüs Ağrısı (atipik)           | -0,641 | 0,699 | 0,840  | 1 | 0,359  | 0,527  | 0,134 | 2,074  |
|         | Göğüs Ağrısı (anjin dışı)       | 0,117  | 0,622 | 0,035  | 1 | 0,851  | 1,124  | 0,332 | 3,807  |
|         | ST Çökmesi                      | -0,860 | 0,183 | 22,156 | 1 | <0,001 | 0,423  | 0,296 | 0,605  |
|         | Damar Sayısı                    | -0,753 | 0,172 | 19,287 | 1 | <0,001 | 0,471  | 0,336 | 0,659  |
|         | ST Kusur                        |        |       | 44,052 | 2 | <0,001 |        |       |        |
|         | ST Kusur (Sabit)                | 1,518  | 0,710 | 4,575  | 1 | 0,032  | 4,565  | 1,136 | 18,349 |
|         | ST Kusur (tersine çevrilebilir) | 2,812  | 0,424 | 43,952 | 1 | <0,001 | 16,648 | 7,249 | 38,232 |
|         | Sabit                           | 4,389  | 0,898 | 23,868 | 1 | <0,001 | 80,550 |       |        |
| Aşama 6 | Cinsiyet (Erkek)                | 1,559  | 0,428 | 13,294 | 1 | <0,001 | 4,756  | 2,057 | 10,998 |
|         | Göğüs Ağrısı                    |        |       | 38,652 | 3 | <0,001 |        |       |        |
|         | Göğüs Ağrısı (tipik)            | -2,370 | 0,615 | 14,836 | 1 | <0,001 | 0,094  | 0,028 | 0,312  |
|         | Göğüs Ağrısı (atipik)           | -0,913 | 0,734 | 1,549  | 1 | 0,213  | 0,401  | 0,095 | 1,691  |
|         | Göğüs Ağrısı (anjin dışı)       | 0,060  | 0,639 | 0,009  | 1 | 0,925  | 1,062  | 0,304 | 3,714  |
|         | ST Çökmesi                      | -0,599 | 0,205 | 8,504  | 1 | 0,004  | 0,549  | 0,367 | 0,822  |
|         | ST Eğim                         |        |       | 14,177 | 2 | 0,001  |        |       |        |
|         | ST Eğim (yukarı doğru)          | -0,482 | 0,804 | 0,360  | 1 | 0,548  | 0,617  | 0,128 | 2,983  |
|         | ST Eğim (Düz)                   | -1,500 | 0,411 | 13,316 | 1 | <0,001 | 0,223  | 0,100 | 0,499  |
|         | Damar Sayısı                    | -0,820 | 0,187 | 19,229 | 1 | <0,001 | 0,440  | 0,305 | 0,635  |
|         | ST Kusur                        |        |       | 39,611 | 2 | <0,001 |        |       |        |
|         | ST Kusur (Sabit)                | 1,496  | 0,736 | 4,135  | 1 | 0,042  | 4,464  | 1,056 | 18,876 |

**Tablo 4.21.** İleriye Dönük Model İçin Değişkenlerin Anlamlılık Değerleri (devam)

|         |                                 |        |       |        |   |        |         |       |        |
|---------|---------------------------------|--------|-------|--------|---|--------|---------|-------|--------|
|         | ST Kusur (tersine çevrilebilir) | 2,732  | 0,435 | 39,530 | 1 | <0,001 | 15,367  | 6,557 | 36,015 |
|         | Sabit                           | 5,022  | 0,959 | 27,420 | 1 | <0,001 | 151,672 |       |        |
| Aşama 7 | Cinsiyet (Erkek)                | 1,523  | 0,435 | 12,243 | 1 | <0,001 | 4,585   | 1,954 | 10,759 |
|         | Göğüs Ağrısı                    |        |       | 26,440 | 3 | <0,001 |         |       |        |
|         | Göğüs Ağrısı (tipik)            | -2,058 | 0,622 | 10,938 | 1 | 0,001  | 0,128   | 0,038 | 0,432  |
|         | Göğüs Ağrısı (atipik)           | -0,948 | 0,736 | 1,660  | 1 | 0,198  | 0,388   | 0,092 | 1,639  |
|         | Göğüs Ağrısı (anjin dışı)       | 0,082  | 0,638 | 0,016  | 1 | 0,898  | 1,085   | 0,311 | 3,786  |
|         | EBA                             | 1,000  | 0,396 | 6,363  | 1 | 0,012  | 2,717   | 1,250 | 5,909  |
|         | ST Çökmesi                      | -0,578 | 0,213 | 7,365  | 1 | 0,007  | 0,561   | 0,370 | 0,852  |
|         | ST Eğim                         |        |       | 11,111 | 2 | 0,004  |         |       |        |
|         | ST Eğim (Yukarı doğru)          | -0,383 | 0,844 | 0,206  | 1 | 0,650  | 0,682   | 0,130 | 3,566  |
|         | ST Eğim (Düz)                   | -1,340 | 0,418 | 10,295 | 1 | 0,001  | 0,262   | 0,116 | 0,594  |
|         | Damar Sayısı                    | -0,836 | 0,189 | 19,595 | 1 | <0,001 | 0,433   | 0,299 | 0,628  |
|         | ST Kusur                        |        |       | 38,793 | 2 | <0,001 |         |       |        |
|         | ST Kusur (Sabit)                | 1,673  | 0,751 | 4,965  | 1 | 0,026  | 5,327   | 1,223 | 23,197 |
|         | ST Kusur (tersine çevrilebilir) | 2,798  | 0,449 | 38,783 | 1 | <0,001 | 16,418  | 6,805 | 39,611 |
|         | Sabit                           | 4,027  | 1,019 | 15,623 | 1 | <0,001 | 56,109  |       |        |

**Tablo 4.22.** İleriye Dönük Modelden Anlamlı Değişkenler Çıkarıldığında Elde Edilen Değişimler

| Değişken Modelden Çıkarıldığında |              |          |                    |    |                     |
|----------------------------------|--------------|----------|--------------------|----|---------------------|
| Değişken                         |              | Model LL | -2 LL deki Değişim | sd | Değişim Anlamlılığı |
| Aşama 1                          | Göğüs Ağrısı | -208,819 | 85,941             | 3  | $<0,001$            |
| Aşama 2                          | Göğüs Ağrısı | -177,499 | 62,020             | 3  | $<0,001$            |
|                                  | ST Çökmesi   | -165,848 | 38,719             | 1  | $<0,001$            |
| Aşama 3                          | Göğüs Ağrısı | -159,655 | 54,976             | 3  | $<0,001$            |
|                                  | ST Çökmesi   | -148,994 | 33,654             | 1  | $<0,001$            |
|                                  | Damar Sayısı | -146,489 | 28,645             | 1  | $<0,001$            |
| Aşama 4                          | Göğüs Ağrısı | -148,992 | 48,816             | 3  | $<0,001$            |
|                                  | ST Çökmesi   | -138,998 | 28,827             | 1  | $<0,001$            |
|                                  | Damar Sayısı | -137,376 | 25,584             | 1  | $<0,001$            |
|                                  | ST Kusur     | -132,167 | 15,164             | 1  | $<0,001$            |
| Aşama 5                          | Cinsiyet     | -124,584 | 10,948             | 1  | $0,001$             |
|                                  | Göğüs Ağrısı | -143,777 | 49,332             | 3  | $<0,001$            |
|                                  | ST Çökmesi   | -132,615 | 27,009             | 1  | $<0,001$            |
|                                  | Damar Sayısı | -129,938 | 21,656             | 1  | $<0,001$            |
|                                  | ST Kusur     | -124,642 | 11,064             | 1  | $0,001$             |
| Aşama 6                          | Cinsiyet     | -119,027 | 15,039             | 1  | $<0,001$            |
|                                  | Göğüs Ağrısı | -134,502 | 45,988             | 3  | $<0,001$            |
|                                  | ST Çökmesi   | -116,224 | 9,431              | 1  | $0,002$             |
|                                  | Eğim         | -119,111 | 15,205             | 2  | $<0,001$            |
|                                  | Damar Sayısı | -122,448 | 21,880             | 1  | $<0,001$            |
|                                  | ST Kusur     | -117,298 | 11,580             | 1  | $0,001$             |
| Aşama 7                          | Cinsiyet     | -115,236 | 13,828             | 1  | $<0,001$            |
|                                  | Göğüs Ağrısı | -123,228 | 29,813             | 3  | $<0,001$            |
|                                  | EBA          | -111,508 | 6,372              | 1  | $0,012$             |
|                                  | ST Çökmesi   | -112,346 | 8,049              | 1  | $0,005$             |
|                                  | Eğim         | -114,153 | 11,662             | 2  | $0,003$             |
|                                  | Damar Sayısı | -119,630 | 22,617             | 1  | $0,000$             |
|                                  | ST Kusur     | -113,430 | 10,217             | 1  | $0,001$             |

**Tablo 4.23. Modele Dâhil Edilmeyen Değişkenler**

| <b>Modelde Olmayan Değişkenler</b> |                         |                                           |             |           |          |
|------------------------------------|-------------------------|-------------------------------------------|-------------|-----------|----------|
| <b>Aşama 1 ..... Aşama 6</b>       |                         |                                           |             |           |          |
|                                    |                         |                                           | <b>Skor</b> | <b>sd</b> | <b>p</b> |
| <b>Aşama 7</b>                     | <b>Değişkenler</b>      | Yaş                                       | 1,918       | 1         | 0,166    |
|                                    |                         | Dinlenme KB                               | 3,372       | 1         | 0,066    |
|                                    |                         | Kolesterol                                | 1,781       | 1         | 0,182    |
|                                    |                         | AKŞ (<120 mg/dL)                          | 0,127       | 1         | 0,722    |
|                                    |                         | Dinlenme EKG                              | 3,075       | 2         | 0,215    |
|                                    |                         | Dinlenme EKG (ST anormallığı)             | 2,709       | 1         | 0,100    |
|                                    |                         | Dinlenme EKG (sol ventrikül hipertrofisi) | 2,988       | 1         | 0,084    |
|                                    | <b>Genel İstatistik</b> |                                           | 10,285      | 7         | 0,173    |

Ancak model incelemesi burada henüz sona ermemiştir. Artıkların incelenmesi ve etkin veya etkili gözlem varsa belirlenmesi gerekmektedir. Bu amaçla, modeli oluştururken tahmin edilen y değeri, tahmin edilen kategori, lojit artıklar, sapma değerleri, Cook's uzaklıkları, kaldıraç (leverage) noktaları ve DFBETA değerlerinin veri kümesine kaydedilmesi sağlanır. Böylece, hangi gözleme ait artık değeri veya uzaklık değerinin model yapısını olumsuz etkilediği anlaşılır.

Lojit artık değerlerinin en küçüğü -29,77 ve en büyüğü 32,79 olarak hesaplanmıştır. Bu değerler oldukça büyük ve sapma değerleri yüksek gözlemlerdir. Sapma değerlerinin negatif veya pozitif yönlü olarak sıfıra yakın olanları tercih edileceğinden büyük sapma değerlerine sahip olan gözlemlerin incelenmesi gereklidir. Ayrıca Cook's uzaklıkları incelendiğinde bazı gözlemlere ait uzaklıkların 1'e yakın olduğu görülmüştür. Etkin gözlemler ile incelenen ölçütlerin genellikle hepsi birbiriyle ilişkilidir. Sapma değerleri yüksek olan gözlemlerin aynı zamanda Cook's uzaklığı ve kaldıraç değerlerinin de normal aralık sınırlarının dışında olduğu görülmüştür. Bu şartlara uyan 8 tane gözlemin modelden çıkarılmasına karar verilmiştir. Çünkü toplam gözlem sayısı 303 olduğu ve çıkarılması gereken gözlem sayısı  $8/303=2,6\%$  olduğu için genel olarak model yapısını bozacak bir oran olmadığı görülmektedir. Bu gözlemler çıkarıldıktan sonra ileriye dönük model seçim yöntemi ile yeniden düzenlenen modelin Omnibus ki-kare değeri 248,52 ( $p<0,001$ ) ve model uyum değerleri  $-2LL=157,96$ ,  $R^2_{Cox\&Snell}=0,569$  ve  $R^2_{Nagelkerke}=0,761$  bulunmuştur.

Hosmer & Lemeshow ki-kare=10,348 ve p=0,241 olduğundan model anlamlılığı mevcuttur. 7 aşamalı modelde sınıflandırma tablosuna göre tahmin başarı oranı %86,8 bulunmuştur. Tüm bu değerleri oluşturulan modelin çok daha iyi olduğunu göstermektedir. Modele katkı yapan değişkenlerin anlamlılık düzeyleri daha da yükselmiştir (hata değerleri daha da küçülmüştür).

Yukarıdaki modelden farklı olarak atipik göğüs ağrısı kategorisinin de modele anlamlı katkı yaptığı görülmüştür. Erkek cinsiyet ve egzersiz esnasında mevcut olan göğüs ağrısı modele pozitif katkı yaparken diğer anlamlı değişkenlerin tamamı negatif katkı yapmaktadır (Tablo 4.20).

**Tablo 4.24.** Etkin Gözlemlerin Çıkarılmasıyla Elde Edilen Model

| Modeldeki Değişkenler |                            |        |       |        |    |        |                  |                              |        |
|-----------------------|----------------------------|--------|-------|--------|----|--------|------------------|------------------------------|--------|
|                       |                            | Beta   | SH    | Wald   | sd | p      | Exp(Beta)<br>=OR | OR için %95<br>Güven Aralığı |        |
|                       |                            |        |       |        |    |        |                  | Alt                          | Üst    |
| Aşama 7               | Cinsiyet (Erkek)           | 2,931  | 0,707 | 17,199 | 1  | <0,001 | 18,743           | 4,691                        | 74,886 |
|                       | Göğüs Ağrısı               |        |       | 30,561 | 3  | <0,001 |                  |                              |        |
|                       | Göğüs Ağrısı (tipik)       | -4,554 | 0,976 | 21,756 | 1  | <0,001 | 0,011            | 0,002                        | 0,071  |
|                       | Göğüs Ağrısı (atipik)      | -3,662 | 1,094 | 11,211 | 1  | 0,001  | 0,026            | 0,003                        | 0,219  |
|                       | Göğüs Ağrısı (anjini dışı) | -0,049 | 0,809 | 0,004  | 1  | 0,952  | 0,952            | 0,195                        | 4,654  |
|                       | EBA                        | 1,415  | 0,552 | 6,568  | 1  | 0,010  | 4,117            | 1,395                        | 12,150 |
|                       | ST Çökmesi                 | -1,154 | 0,345 | 11,206 | 1  | 0,001  | 0,315            | 0,160                        | 0,620  |
|                       | ST Eğim                    |        |       | 21,815 | 2  | <0,001 |                  |                              |        |

**Tablo 4.25.** Etkin Gözlemlerin Çıkarılmasıyla Elde Edilen Model (devam)

|                           |            |       |        |   |        |          |       |       |
|---------------------------|------------|-------|--------|---|--------|----------|-------|-------|
| ST Eğim<br>(Yukarı doğru) | -<br>0,762 | 1,254 | 0,370  | 1 | 0,543  | 0,467    | 0,040 | 5,446 |
| ST Eğim<br>(Düz)          | -<br>3,418 | 0,746 | 21,021 | 1 | <0,001 | 0,033    | 0,008 | 0,141 |
| Damar<br>Sayısı           | -<br>2,143 | 0,396 | 29,305 | 1 | <0,001 | 0,117    | 0,054 | 0,255 |
| ST<br>Kusur               | -<br>2,015 | 0,432 | 21,736 | 1 | <0,001 | 0,133    | 0,057 | 0,311 |
| Sabit                     | 9,815      | 1,848 | 28,209 | 1 | <0,001 | 18310,46 |       |       |

### Çok Kategorili (Multinomial) Lojistik Regresyon Modeli

Bu çalışmanın veri kümesi, sağlık sigortası ödemesi yapan müşterilere ait poliçe bilgilerinden elde edilmiştir. Toplam gözlem sayısı 1338 müşteriye ait olup erkek ve kadın müşteri oranları birbirine oldukça yakındır. Erkek müşterilerin oranları %50,52 ve kadınların oranı %49,48 olarak verilmiştir (Şekil 4.4).

Ülkenin güneydoğu eyaletlerinde olan müşterilerin oranı biraz daha yüksek (%27,2), diğer bölgeler ait oranlar birbirine oldukça yakındır (Şekil 4.5).

Müşterilerin büyük bir bölümü sigara kullanmamaktadır (%79,52) (Şekil 4.6). Müşterilerin genel yaş ortalaması 39,20±14,04 yıl ve çocuk sayısı ortalaması 1,09±1,20 (medyan 1) adet bulunmuştur. Müşterilerin VKİ ortalaması genel olarak yüksek olup 30,66±6,09 kg/m<sup>2</sup> olarak bulunmuştur.

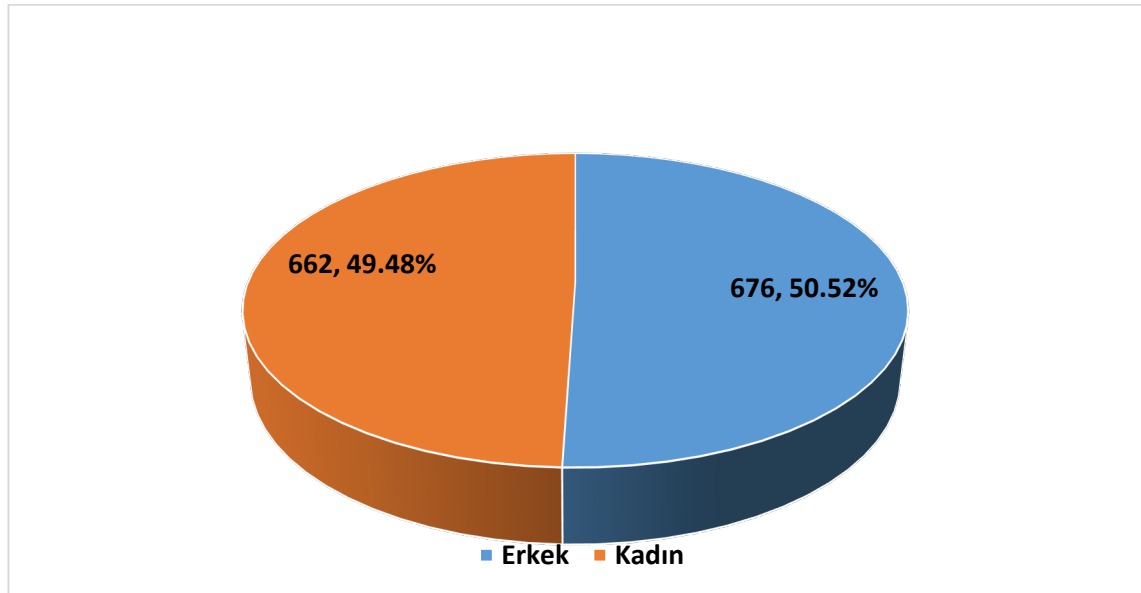
Müşterilerin ödemiş olduğu yıllık sağlık sigortası bedeli ortalama 13541±12315 (medyan 9575) USD olarak hesaplanmıştır. İkamet edilen bölgelere göre müşterilerin özellikleri tek yönlü varyans analizi ile karşılaştırılmıştır.

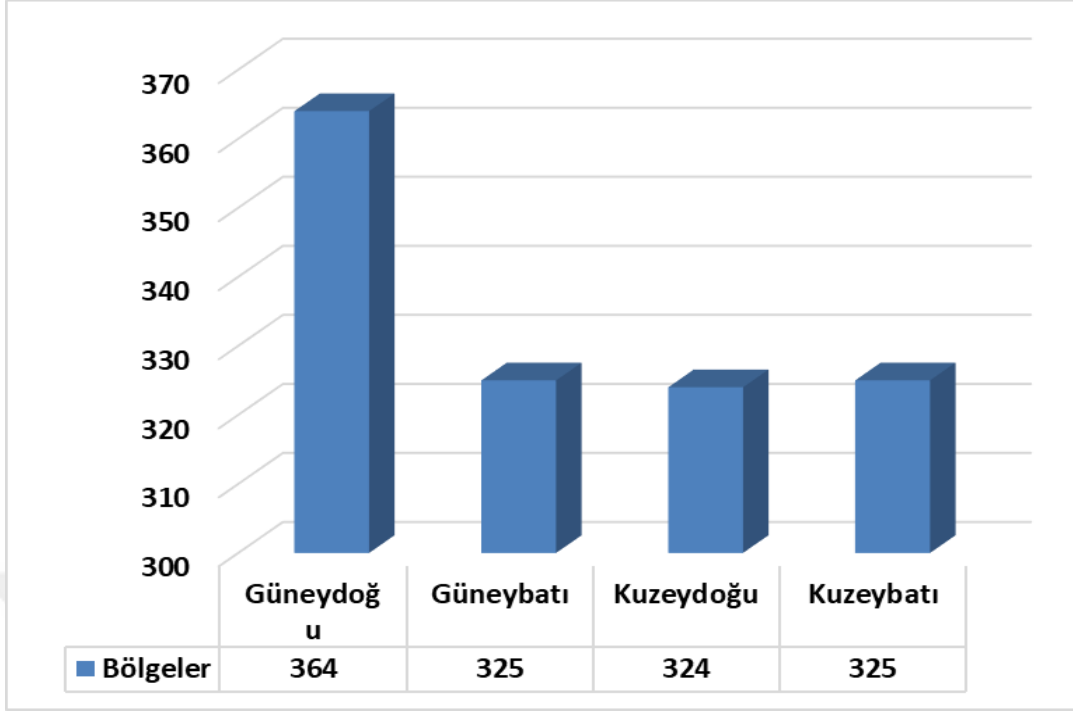
Müşterilerin yaşları ve çocuk sayıları bölgeler arasında anlamlı farklılığa sahip bulunmamıştır. Ancak, müşterilerin VKİ değerleri güneydoğu bölgesinde yaşayanlar için anlamlı düzeyde yüksektir (p<0,001). Müşterilerin ödediği sağlık sigorta miktarı da güneydoğu bölgesi müşterilerinde anlamlı düzeyde yüksek bulunmuştur (p=0,010).

Müşterilerin cinsiyetleri ve sigara kullanım oranları eyaletler arasında anlamlı farklılığa sahip bulunmamıştır (Tablo 4.21).

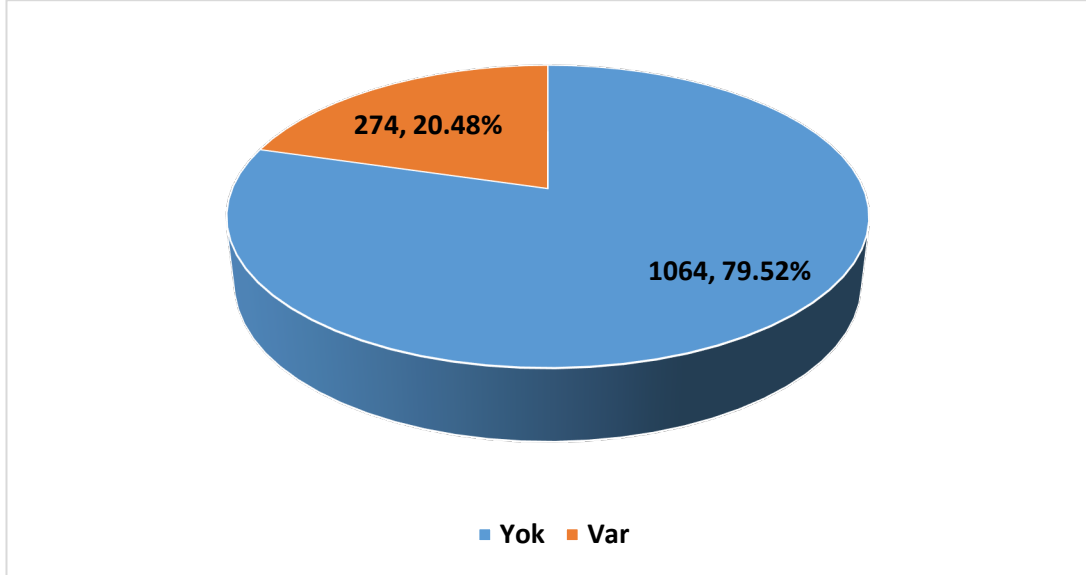
**Tablo 4.26.** Çalışmaya Alınan Müşterilerin Bölgelere Göre Genel Özellikleri

|                             | Güneydoğu<br>(n=364)  | Güneybatı<br>(n=325)  | Kuzeydoğu<br>(n=324)   | Kuzeybatı<br>(n=325)  |          |
|-----------------------------|-----------------------|-----------------------|------------------------|-----------------------|----------|
|                             | Ortalama±SS (Medyan)  |                       |                        |                       | <i>p</i> |
| Yaş (yıl)                   | 38,94±14,16           | 39,46±13,96           | 39,27±14,07            | 39,2±14,05            | 0,971    |
| VKİ (kg/m <sup>2</sup> )    | 33,36±6,48            | 30,6±5,69             | 29,18±5,94             | 29,2±5,14             | <0,001   |
| Çocuk                       | 1,05±1,18             | 1,14±1,28             | 1,05±1,20              | 1,15±1,17             | 0,542    |
| Sigorta<br>Ödemesi<br>(USD) | 14735±13971<br>(9294) | 12346±11557<br>(8798) | 14526±12091<br>(11204) | 12417±11072<br>(8965) | 0,010    |
|                             | N (%)                 | N (%)                 | N (%)                  | N (%)                 |          |
| Cinsiyet                    |                       |                       |                        |                       |          |
| Erkek                       | 189 (51,9)            | 163 (50,2)            | 163 (50,3)             | 161 (49,5)            | 0,933    |
| Kadın                       | 175 (48,1)            | 162 (49,8)            | 161 (49,7)             | 164 (50,5)            |          |
| Sigara                      |                       |                       |                        |                       |          |
| Kullanmıyor                 | 273 (75,0)            | 267 (82,2)            | 257 (79,3)             | 267 (82,2)            | 0,062    |
| Kullanıyor                  | 91 (25,0)             | 28(17,8)              | 67 (20,7)              | 58 (17,8)             |          |

**Şekil 4.4.** Sağlık Sigortası Ödemesi Yapan Müşterilerin Cinsiyet Oranları



**Şekil 4.5.** Sağlık Sigortası Ödemesi Yapan Müşterilerin Yaşadıkları Bölgelerin Oranları



**Şekil 4.6.** Sağlık Sigortası Ödemesi Yapan Müşterilerin Sigara Kullanım Oranları

Müşterilerin yaşadıkları bölgeler bağımlı değişken olarak alınarak diğer değişkenlerin yaşanılan bölge üzerindeki etkileri belirlenmeye çalışılmıştır. Bağımlı değişkene ait kategori sayısı 4 olup, “Kuzeybatı” kategorisi veya yapılan kodlamaya göre 4.kategori referans kategori olarak belirlenmiştir. Bu seçimin belirgin bir özelliği

yoktur. Rastgele bir seçim nedeniyle belirlenmiştir. Genellikle birbiri ile ilişkili veya ters yönlü kategoriler olduğunda belirgin referans kategori seçimi yapılır.

Çok kategorili lojistik regresyon modeli iki kategorili lojistik regresyonu modelinden biraz farklı bir yapıdadır. Çözüm tekniği ve sonuçların yorumlanması benzerdir. Modelin anlamlılığı, modelin uyumu veya katsayıların anlamlılığı için yapılan analizler ve kullanılan uyum iyiliği testleri benzerdir. Temel fark modelin kuruluş yöntemidir. Bağımlı değişkene ait her bir kategori için ayrı ayrı koşullu olasılıklar oluşturulur ve tüm olasılıklar toplanarak genel bir lojit modeli oluşturulur.

İki kategorili lojistik modelde olduğu gibi öncelikle modele yalnızca sabit katsayının eklendiği model ile başlanır. Modelin Akaike Information Criterion (AIC) ve Bayesian Information Criterion (BIC) değerleri sırasıyla 3602,95 ve 3712,13'e gerilemiştir. Buna göre -2LL değeri 3560,95 hesaplanmıştır. Modelin anlamlılık ki-kare istatistiği 145,35 ve  $p < 0,001$  bulunduğundan oluşturulan model anlamlıdır (Tablo 4.22). Yani  $\beta_0$  katsayısı anlamlı bulunmuş olup “sıfır” değildir.

Modele ilişkin uyum iyiliği testleri Pearson ve Deviance (Sapma) istatistikleri kullanılarak yapılır. İstatistikler incelendiğinde modelin uyum düzeyinin anlamlı olmadığı görülmüştür. Çünkü hem Pearson hem de Sapma ki-kare istatistikleri anlamlı bulunmamış ( $p > 0,05$ ), sıfır hipotezi reddedilememiştir (Tablo 4.23). Ancak modele bağımsız değişkenler henüz eklenmediğinden yorum yapmak için erkendir. Sonrasında modele değişkenlerin eklenmesine geçilebilir. Model “tam” model olarak, yani “Enter” seçim yöntemi ile oluşturulmuştur. Modele ilişkin  $R^2$  değerleri düşük bulunmuştur. Cox&Snell  $R^2 = 0,103$  ve Nagelkerke  $R^2 = 0,110$  elde edilmiştir. Bu durum modelin açıklayıcılık düzeyinin yeterince yüksek olmadığını göstermektedir (Tablo 4.24).

Bağımsız değişkenler eklenince modelin uyumunun ve anlamlılığının hangi düzeyde olduğunu belirlemek için Olabilirlik Oran Testi sonuçları incelenir. Her bir değişken için AIC, BIC ve -2LL değerleri hesaplanır. Değerler birbirine oldukça yakın bulunmuştur ve bu kriter göstergelerinde çıkarım yapmak mümkün değildir. Ancak Olabilirlik Oran Testi sonuçları hangi değişkenin modele katkı yapabileceği konusunda fikir vermektedir.

Müşterilere ait VKİ, sigorta ücreti ve sigara kullanımı ki-kare istatistiklerinin anlamlı bulunması nedeniyle bu değişkenlerin modelde anlamlı olabileceği düşünülür (Tablo 4.25).

**Tablo 4.27. Çok Kategorili Lojistik Modele İlişkin Uyum Bilgileri**

| <b>Model Uyum Bilgileri</b> |                              |            |                               |                |           |          |
|-----------------------------|------------------------------|------------|-------------------------------|----------------|-----------|----------|
| <b>Model</b>                | <b>Model Uyum Kriterleri</b> |            | <b>Olabilirlik Oran Testi</b> |                |           |          |
|                             | <b>AIC</b>                   | <b>BIC</b> | <b>-2 Log Likelihood</b>      | <b>Ki-kare</b> | <b>sd</b> | <b>p</b> |
| Sabit Model                 | 3712,317                     | 3727,914   | 3706,317                      |                |           |          |
| Final Model                 | 3602,958                     | 3712,136   | 3560,958                      | 145,359        | 18        | <0,001   |

**Tablo 4.28. Çok Kategorili Lojistik Modele İlişkin Uyum İyiliği Sonuçları**

| <b>Uyum iyiliği</b> |                |           |          |
|---------------------|----------------|-----------|----------|
|                     | <b>Ki-kare</b> | <b>sd</b> | <b>p</b> |
| Pearson             | 4018,810       | 3990      | 0,371    |
| Sapma               | 3560,958       | 3990      | 1,000    |

**Tablo 4.29. Çok Kategorili Modele İlişkin R<sup>2</sup> Değerleri**

| <b>Başlangıç R<sup>2</sup></b> |       |
|--------------------------------|-------|
| Cox & Snell                    | 0,103 |
| Nagelkerke                     | 0,110 |
| McFadden                       | 0,039 |

**Tablo 4.30. Olabilirlik Oran Testi Sonuçları**

| <b>Olabilirlik Oran Testi</b> |                              |                              |                               |                               |           |          |
|-------------------------------|------------------------------|------------------------------|-------------------------------|-------------------------------|-----------|----------|
| <b>Etken</b>                  | <b>Model Uyum Kriterleri</b> |                              |                               | <b>Olabilirlik Oran Testi</b> |           |          |
|                               | <b>AIC İndirgenmiş Model</b> | <b>BIC İndirgenmiş Model</b> | <b>İndirgenmiş Model -2LL</b> | <b>Ki-kare</b>                | <b>sd</b> | <b>p</b> |
| Sabit                         | 3602,958                     | 3712,136                     | 3560,958                      | 0,000                         | 0         |          |
| Yaş                           | 3599,342                     | 3692,923                     | 3563,342                      | 2,384                         | 3         | 0,497    |
| VKİ                           | 3721,298                     | 3814,879                     | 3685,298                      | 124,340                       | 3         | <0,001   |
| Çocuk Sayısı                  | 3599,701                     | 3693,281                     | 3563,701                      | 2,742                         | 3         | 0,433    |
| Sigorta Ücreti                | 3618,869                     | 3712,449                     | 3582,869                      | 21,910                        | 3         | <0,001   |

**Tablo 4.31. Olabilirlik Oran Testi Sonuçları (devam)**

|          |          |          |          |        |   |       |
|----------|----------|----------|----------|--------|---|-------|
| Cinsiyet | 3596,966 | 3690,547 | 3560,966 | 0,008  | 3 | 1,000 |
| Sigara   | 3614,626 | 3708,207 | 3578,626 | 17,668 | 3 | 0,001 |

Bağımlı değişken dört kategorili olup “Kuzeybatı” kategorisi referans olarak belirlenmiştir. Bu nedenle oluşturulan modelde “Kuzeybatı” kategorisine ilişkin sonuçlar bulunmaz. Diğer bölgelere ait katsayı tahminleri, anlamlılık değerleri ve OR değerleri Tablo 4.26’da sunulmuştur. Her bir kategori için bulunan sonuçlar referans alınan kategoriye göre yorumlanır.

İki kategorili lojistik regresyon modelinde olduğu gibi Beta katsayılarının anlamlılık sınaması Wald istatistiği ile gerçekleştirilir.

Güneydoğu bölgesinde yaşayan müşteriler için sabit katsayı anlamlı bulunmuştur. Bunun yanı sıra, müşterilerin VKİ değerleri Kuzeybatı kategorisine göre anlamı daha yüksektir. Çünkü Beta=0,129, Wald=84,84 ve  $p<0,001$  bulunmuştur. Beta katsayısına göre  $\text{Exp}(\text{Beta})=1,137$  ve %95 güven aralığı 1,106-1,169 “1” değerini içermemektedir. Yani, VKİ değerleri Güneydoğu bölgesinde yaşıyor olma olasılığını Kuzeybatı bölgesinde yaşayan müşterilere göre 1,13 kat artırmaktadır.

Çocuk sayısı, sigorta ücreti ve cinsiyetin Güneydoğu bölgesinde yaşama olasılığı üzerinde anlamlı etkisi bulunmamıştır. Ancak sigara kullanmıyor olma durumunun Güneydoğu bölgesinde yaşıyor olma olasılığını azaltıcı etkisi bulunmuştur. Beta=-0,950 ( $p=0,006$ ) ve OR=0,387 (0,197-0,758) bulunduğundan  $1/0,387=2,583$  olarak hesaplanır. Yani, Kuzeybatı bölgesinde yaşayan müşterilere göre Güneydoğu bölgesinde yaşayan müşterilerin sigara içmeme oranı 2,58 kat daha yüksektir yorumu yapılmıştır.

Güneybatı bölgesi için yalnızca VKİ değerinin pozitif yönlü bir etkisinin mevcut olduğu görülmüştür. Beta=0,047 ( $p=0,001$ ) ve OR=1,049 hesaplandığından oldukça düşük bir artırıcı etki gözlenmiştir.

Kuzeybatı bölgesinde yalnızca ödemesi yapılan sağlık sigortası ücreti için anlamlı Wald istatistiği bulunmaktadır. Ancak, Beta katsayısı 0,0 olarak hesaplandığından OR değeri 1,0 bulunacaktır. Bu durum, zaman zaman modellerde karşımıza çıkabilmektedir.

Wald istatistiğinin anlamlı bulunması o değişkenin anlamlı bir risk veya koruyucu faktör olabileceği anlamına gelmemektedir. Burada da böyle bir durum oluşmuştur. OR değeri 1 olduğunda o değişken için herhangi bir yorum yapmak mümkün değildir. Bu durumun çeşitli nedenleri olabilir. Örneklemeye alınan gözlemlerin özelliği, sayısı, ölçülen değişkenin yapısı gibi durumlar bunun nedeni olmuş olabilir. Yani, sağlık sigortası ödeme tutarları arasında kuzeydoğu ve kuzeybatı bölgeleri arasında benzerlik bulunmaktadır.

Her bir kategori için yapılan tahminlerin başarısı sınıflandırma tablosunda gösterilmiştir. En iyi tahmin oranları Güneydoğu bölgesine ait olanlardır (%59,1). Diğer bölgeler için yapılan tahminlerin başarı oranları oldukça düşük ve genel olarak modelin genel başarı oranı %32,7 bulunmuştur (Tablo 4.27).

**Tablo 4.32. Modele İlişkin Katsayı Tahminleri ve Anlamlılık Düzeyleri**

| Parametre Tahminleri |                    |        |       |        |    |          |       |                           |       |
|----------------------|--------------------|--------|-------|--------|----|----------|-------|---------------------------|-------|
| Bölge                |                    | Beta   | SH    | Wald   | sd | <i>p</i> | OR    | OR için %95 Güven Aralığı |       |
|                      |                    |        |       |        |    |          |       | Alt                       | Üst   |
| Güneydoğu            | Sabit              | -2,758 | 0,485 | 32,298 | 1  | <0,001   |       |                           |       |
|                      | Yaş                | -0,002 | 0,006 | 0,067  | 1  | 0,796    | 0,998 | 0,986                     | 1,011 |
|                      | VKİ                | 0,129  | 0,014 | 84,848 | 1  | <0,001   | 1,137 | 1,106                     | 1,169 |
|                      | Çocuk Sayısı       | -0,066 | 0,062 | 1,108  | 1  | 0,292    | 0,936 | 0,829                     | 1,058 |
|                      | Sigorta Ücreti     | 0,000  | 0,000 | 2,483  | 1  | 0,115    | 1,000 | 1,000                     | 1,000 |
|                      | [Cinsiye t =Erkek] | 0,000  | 0,150 | 0,000  | 1  | 0,999    | 1,000 | 0,745                     | 1,342 |
|                      | [Cinsiye t =Kadın] | 0      |       |        | 0  |          |       |                           |       |
|                      | [Sigara=Yok]       | -0,950 | 0,343 | 7,668  | 1  | 0,006    | 0,387 | 0,197                     | 0,758 |
|                      | [Sigara=Var]       | 0      |       |        | 0  |          |       |                           |       |
| Güneybatı            | Sabit              | -1,045 | 0,480 | 4,753  | 1  | 0,029    |       |                           |       |
|                      | Yaş                | 0,003  | 0,006 | 0,220  | 1  | 0,639    | 1,003 | 0,991                     | 1,015 |
|                      | VKİ                | 0,047  | 0,014 | 11,970 | 1  | 0,001    | 1,049 | 1,021                     | 1,077 |
|                      | Çocuk Sayısı       | 0,000  | 0,061 | 0,000  | 1  | 0,995    | 1,000 | 0,888                     | 1,127 |

**Tablo 4.33.** Modele İlişkin Katsayı Tahminleri ve Anlamlılık Düzeyleri (devam)

|           |                  |        |       |       |   |       |       |       |       |
|-----------|------------------|--------|-------|-------|---|-------|-------|-------|-------|
|           | Sigorta Ücreti   | 0,000  | 0,000 | 1,551 | 1 | 0,213 | 1,000 | 1,000 | 1,000 |
|           | [Cinsiyet=Erkek] | 0,006  | 0,149 | 0,001 | 1 | 0,970 | 1,006 | 0,750 | 1,348 |
|           | [Cinsiyet=Kadın] | 0      |       |       | 0 |       |       |       |       |
|           | [Sigara=Yok]     | -0,362 | 0,334 | 1,176 | 1 | 0,278 | 0,696 | 0,362 | 1,340 |
|           | [Sigara=Var]     | 0      |       |       | 0 |       |       |       |       |
| Kuzeydoğu | Sabit            | -0,103 | 0,478 | 0,047 | 1 | 0,829 |       |       |       |
|           | Yaş              | -0,007 | 0,006 | 1,249 | 1 | 0,264 | 0,993 | 0,982 | 1,005 |
|           | VKİ              | -0,011 | 0,014 | 0,685 | 1 | 0,408 | 0,989 | 0,962 | 1,016 |
|           | Çocuk Sayısı     | -0,085 | 0,062 | 1,868 | 1 | 0,172 | 0,919 | 0,814 | 1,037 |
|           | Sigorta Ücreti   | 0,000  | 0,000 | 7,532 | 1 | 0,006 | 1,000 | 1,000 | 1,000 |
|           | [Cinsiyet=Erkek] | 0,012  | 0,150 | 0,006 | 1 | 0,937 | 1,012 | 0,755 | 1,356 |
|           | [Cinsiyet=Kadın] | 0      |       |       | 0 |       |       |       |       |
|           | [Sigara=Yok]     | 0,472  | 0,305 | 2,394 | 1 | 0,122 | 1,603 | 0,882 | 2,913 |
|           | [Sigara=Var]     | 0      |       |       | 0 |       |       |       |       |

**Tablo 4.34.** Model Başarı Oranını Gösteren Sınıflandırma Tablosu

| Sınıflandırma Tablosu |               |           |           |           |                         |
|-----------------------|---------------|-----------|-----------|-----------|-------------------------|
| Gözlenen              | Tahmin Edilen |           |           |           |                         |
|                       | Güneydoğu     | Güneybatı | Kuzeydoğu | Kuzeybatı | Düzeltilmiş Yüzde Oranı |
| Güneydoğu             | 215           | 41        | 38        | 70        | 59,1%                   |
| Güneybatı             | 145           | 41        | 57        | 82        | 12,6%                   |
| Kuzeydoğu             | 112           | 31        | 90        | 91        | 27,8%                   |
| Kuzeybatı             | 117           | 41        | 75        | 92        | 28,3%                   |
| Toplam yüzde Oranı    | 44,0%         | 11,5%     | 19,4%     | 25,0%     | 32,7%                   |

Tam model yerine ileriye dönük aşamalı model kurgulandığından bilgi kriterleri ve anlamlı bulunan değişkenlerde değişiklik gözlenmemiştir. Benzer değişkenler, yani sigara kullanımı ve VKİ değişkenleri anlamlı etkiye sahip bulunmuştur. Benzer şekilde, sigorta ödemesi anlamlı bulunmuş ancak etkili olmayan OR değerine sahip olduğu görülmüştür.

İki kategorili lojistik regresyon modelinde olduğu gibi artıkların incelenmesi gerçekleştirilmiştir. Ancak çıkarılması gereken gözlem değerleri bulunmamıştır.

Gerek uzaklıklar gerekse artıklardan aşırı büyük veya aşırı küçük değerlere rastlanmamıştır. Bu nedenle, gözlemlerin çıkarılması ile anlamlılık düzeyi daha yüksek bir model elde edilememiştir. Ayrıca, VIF ve koşul indeksi değerlerinin incelenmesine göre modelde çoklu bağlantı problemi de bulunmamaktadır.

## 5. TARTIŞMA

Bu çalışma, lojistik regresyon analizinin temellerini açıklamak ve sağlık alanında örnek uygulamalarını göstermek için hazırlanmıştır. Lojistik regresyonun teorik geri planı anlatılmıştır. Lojistik regresyon, doğrusal regresyondan farklı olarak bağımlı değişken kategorik değerler aldığı zaman kullanılır. Kurulan modelde bağımsız değişkenlerin oluşturduğu denklem nümerik değerler alır. Ancak lojit dönüşüm yardımıyla sigmoid fonksiyonuna dönüşür. Böylece 0 ile 1 arasında bir değer alır. Bu değerler 0,5'in altında ise 0, üzerinde ise 1 olarak alınır ve böylece iki kategorili lojistik regresyon analizi tahmin değerleri oluşmuş olur. Lojistik regresyon hemen her alanda yoğun olarak kullanılmaktadır. Sosyal bilimlerden mühendislik bilimlerine, eğitim bilimlerinden sağlık bilimlerine kadar hemen her alanda kullanılan lojistik regresyon modelleri son yıllarda veri bilimi alanında da tahmin modellemesi için gözetimli öğrenme (sınıflama) alanında kullanılmaktadır. Ancak sağlık alanında yapılan uygulamalarda genellikle bağımsız değişkenlerin etki veya risk düzeylerini belirlemek için kullanılmaktadır. Tahmin değerlerinin üstel değerleri, belirlenen faktörler için risk veya koruyuculuk düzeylerini göstermektedir. Genel olarak lojistik regresyon: 1. İkili lojistik regresyon. Kategorik cevap değişkeni mümkün olan 2 sonuç ile ifade edilir. Örneğin hasta veya değil. 2. Çok terimli (multinomial) lojistik regresyon 3 veya daha fazla kategorik değerlerden oluşur. Örneğin yeni doğan bebekler için yapılan ankette sağlıklı, hasta veya ölü doğması durumu. 3. Ordinal lojistik regresyon. 3 veya daha fazla kategorik değerlerin sıralı (ordering) olması ile oluşan regresyon. Örneğin bir hekimin hastalar tarafında değerlendirilmesi 1'den 5'e kadar olan durumu.

Lojistik regresyonda temel amaç, gözlemlerin hangi kategoriye ait olduğunu tahminlemek için bir model oluşturulmasıdır. İlk defa 1800'lü yılların ortalarında sosyo-ekonomik bir konunun araştırılması için kullanıldığı bildirilmektedir. İlerleyen yıllarda alternatif olarak farklı yöntemler önerilmiştir. Hangi yöntemi olursa olsun lojistik regresyonun en önemli avantajı doğrusal modellerde olduğu gibi varsayım gerektirmemesidir. Ayrıca, elde edilen modellerin yorumlanması oldukça kolaydır. Kurulan modellerde bağımsız değişkenler isimsel, sıralı veya oransal ölçekli olabilmektedir. Lojistik regresyon analizinde diğer sınıflama yöntemlerinde farklı

olarak bağlantı (link) fonksiyonu oldukça basittir. Bu nedenle, lojistik regresyon analizinde en çok olabilirlik tahmin edicileri hesaplanırken log odds oranları kullanılır, dağılım kullanılmasına gerek bulunmamaktadır. Bağımlı değişkenin kategorik değerler alması ve bağımsız değişkenlerin normal dağılıma uygun olma zorunluluğunun olmaması değişen varyanslılık sorununun oluşmamasına neden olmaktadır. Lojistik regresyon modellerinde genel olarak otokorelasyon sorunu oluşmamaktadır ancak yine de kontrol edilmesi yerinde olacaktır. Buna karşın, doğrusal regresyon modellerinde de önemli bir sorun olan çoklu bağlantı (multicollinearity) problemi lojistik regresyonda kolaylıkla ortaya çıkabilir.

Lojistik regresyonda genel olarak dört konu başlığına dikkat edilmesi gerekir. Birincisi, model oluşturmak için alınan gözlem sayısının yüksek olması önerilmektedir. İkinci olarak, modelin uyum iyiliği test sonuçlarının mutlaka değerlendirilmesi gerektiğidir. Üçüncü nokta, bağımsız değişkenler arasındaki yüksek korelasyon nedeniyle çoklu bağlantı problemi oluşma ihtimalidir. Dördüncü olarak lojistik modellerin uç değerlere karşı duyarlı olmasıdır. Bu nedenle artıkların (outliers) incelenmesi ve gerekirse uç değerlere neden olan gözlemlerin belirli şartlar altında modelden çıkarılması önerilmektedir.

Çokluk (2010), yukarıdaki bilgileri verdiği çalışmasında eleştirel düşünme düzeyini ölçmek üzere bir Eğitim Fakültesine bağlı Rehberlik ve Psikolojik Danışmanlık Bölümünde öğrenim gören toplam 200 öğrenci ile yapılan korelasyonel anket çalışması sonuçlarını paylaşmıştır. Çalışmada, Bilimsel Düşünme Becerisi Ölçeği, Epistemolojik İnanç Ölçeği ve Fatalizm Ölçeği sonuçları kullanılmıştır. Kurulan lojistik regresyon modelinde çoklu bağlantı olmadığı ve VIF değerinin 1,21' e kadar yükseldiği görülmektedir. Hesaplanan -2LL değerinin 277,23 olduğu ve çok iyi uyum gösterdiği bildirilmiştir. Nagelkerke  $R^2$  değeri 0,135 olarak bulunmuştur. Omnibus testi  $X^2=21,302$  ve  $p<0,001$  değeri bağımlı ve bağımsız değişkenler arasındaki tahmine yönelik ilişkinin varlığını göstermektedir. Adımsal yöntem sonucunda bilimsel düşünce ve epistemolojik inanç değişkenleri anlamlı bulunmuştur (52).

Doğrusal regresyon modellerinde zaman zaman değişen varyanslılık sorunu ortaya çıkmaktadır. Bunun giderilmesi için örneklemin değiştirilmesi genel olarak

mümkün olmadığı için ağırlıklı en küçük kareler yöntemi kullanılmaktadır. Ancak gerçek beklenen değerler ve ağırlıklar bilinmediği için lojit modelin ilişkilerdeki doğrusallığı kurtarabileceği bildirilmiştir. Çalışmada veri kümesi olarak İİBF'den mezun olan 100 öğrenciye ait cinsiyet, mezun olunan lise türü ve üniversite giriş sınavındaki net sayıları bağımlı değişken olarak alınarak dört yıl içerisinde mezun olma durumu bağımlı değişken bilgisi olarak alınmıştır. Yapılan analiz sonucunda modelin anlamlı ve uyumlu olduğu, bağımsız değişkenler içerisinde bitirilen lise türünün fakülteyi dört yılda bitirme ile anlamlı ilişkiye sahip olduğu bildirilmiştir (70).

Regresyon modellerinde en önemlilerinden birisi de model seçimidir. Bu amaçla, çok sayıda paket program lojistik regresyon analizi gerçekleştirmekte ve bunların büyük bir kısmı da adımsal, ileriye dönük veya geriye dönük iteratif yöntem imkânı sunmaktadır. Özellikle, küçük ve orta düzeyde örneklem sayısı mevcutsa modellerde sorun ortaya çıkmaktadır. Bu amaçla bir simülasyon çalışmasında adımsal yöntem ile penalize regresyon yöntemleri olan Ridge, LASO ve Elastic Net yöntemleri karşılaştırılmıştır. Bu çalışmada olduğu gibi, simülasyon çalışması sonrası gerçek veri için açık kaynak kullanıma uygun ve çok sayıda çalışmada kullanılan Pima Indian Diabetes, Ionosphere, Haberman Survival ve Wisconsin tanısal göğüs kanser verileri kullanılmıştır. Gerek simülasyon gerekse gerçek veri kümeleri senaryolarında penalize lojistik regresyon yöntemlerinin adımsal regresyona göre üstün olduğu bildirilmiştir. Ayrıca elastic net yönteminin diğer yöntemlere göre daha iyi sonuçlar verdiği gösterilmiştir (71).

Model seçimi konusunda yer alan diğer bir çalışmada ise kadın işgücünde ücretlerin eğitim düzeyinden nasıl etkilendiği araştırılmıştır. Model seçimi için lojit, probit ve tobit yöntemleri kullanılmıştır. Lojit ve probit yöntemler arasında genel olarak fark olmadığı, tobit regresyon ile yanlış tahmin değerleri elde edildiği; modele eklenen bağımsız değişkenlere bağlı olmakla birlikte en çok olabilirlik tahmin yöntemlerinin daha başarılı olduğu ifade edilmiştir (72).

Benzer tarzda yapılan başka bir çalışmada ise Filipin İstatistik Enstitüsü'nden alınan Hanehalkı Gelir ve Harcama Araştırması anket verileri kullanılmıştır. Toplam 60 sorudan oluşan anket 41544 kişiye uygulanmıştır. Ancak model seçimi çalışmasında dokuz değişken kullanılmış ve sağlık harcamaları bağımlı değişken

olarak belirlenmiştir. Çalışmada lojit, probit ve tobit modellerin başarıları araştırılmıştır. Analizler sonucunda benzer modeller elde edilmesine karşın Tobit modelinde marjın etkisi ile AIC ve BIC kriterlerinin daha düşük bulunduğu gözlenmiştir. Lojit model için  $-2LL=12398$ , Omnibus  $X^2=365,71$  ( $p<0,001$ ) ve  $R^2=0,028$  olarak hesaplanmıştır. Lojit ve probit modellerin daha yaygın kullanılmasına karşın tobit modelin daha iyi sonuç verdiği bildirilmektedir (73).

Bir gözlemin birden fazla popülasyondan hangisine atanacağı sınıflama problemi olarak bilinmektedir. Eğer popülasyonlara ait varyans-kovaryans matrisinin homojen olduğu hipotezi kabul edilirse gözlemin atanacağı kümenin belirlenmesinde ayırma analizi tercih edilir. Ancak çok değişkenli normallik varsayımı yerine gelmediğinde veya homojen varyans-kovaryans hipotezi reddedildiğinde varsayım gerektirmeyen lojistik regresyon analizi tercih edilir. Bu amaçla bir çalışmada, görüş mesafesinin 1000 m altında ve üzerinde olması durumları meteorolojik veriler ile açıklanmaya çalışılmıştır. Eskişehir ilinde hava üssü bulunmaktadır ve yılın belirli bir bölümünde hava sisli olmakta ve görüş mesafesi düşmektedir. Modelin anlamlılık değeri  $X^2=292,16$  ( $p<0,001$ ) olarak bulunmuş ve sınıflama tablosunda “sis yok” kategorisi için doğruluk oranı %98,13 ve “sis var” kategorisi için %56,64 ve genel doğruluk oranı %92,68 olarak hesaplanmıştır. Bu oran sınıflamanın oldukça başarılı bir şekilde tahmin edildiğini göstermektedir. Modelde bağımsız değişken olarak alınmış olan sıcaklık, nem, basınç, rüzgâr hızı ve havanın kapalılığının tamamı anlamlı etkiye sahip bulunmuştur. Rüzgâr hızı ve sıcaklık negatif katsayıya sahip bulunmuş ve buna bağlı olarak ta odds değerleri  $<1$  olarak hesaplanmıştır (74).

Coğrafi bilgi sistemleri yardımıyla lojistik regresyon uygulaması çok sayıda çalışmada görülmektedir. Bir çalışmada Kansas’ın (USA) kuzeydoğu bölgesindeki dağlarda toprak kayması tahmin etmek için çok kategorili lojistik regresyon kullanılmıştır. Bağımsız değişkenler olarak eğim, alüvyon, kar birikintisi, kil, kireç taşı gibi jeolojik yapılar alınmıştır. Model için toplam 2.022.861 piksel bilgisi kullanılmıştır. Modelin anlamlılık değerleri yüksek ( $-2LL=75.834$  ve  $X^2=58901$  ( $p<0,001$ )) bulunmuş, modeldeki tüm bağımsız değişkenlerin anlamlı etkiye sahip oldukları bildirilmiştir. Çok kategorili lojistik regresyon analizinin toprak kayması haritalaması için kullanışlı bir yöntem olduğu rapor edilmiştir (75).

Benzer başka bir çalışmada Japonya’da bulunan Kakudo-Yahiko dağının toprak kayması olasılık haritası çıkarılmıştır. Çalışma hem iki kategorili hem de çok kategorili olarak yapılmıştır. Modelde bağımsız değişkenler olarak yön, yükseklik, kaya zemin, ayırt edici özellikler, toprak bilgisi, eğim açısı ve yola yakınlık durumu alınmıştır. İki kategorili model oldukça anlamlı bulunmuştur.  $-2LL=447.111$ , Omnibus  $X^2=87.306$  ve  $R^2=0,195$  olarak hesaplanmıştır. Toplam 10.584.768 piksele ait değer ile model oluşturulmuş olup sınıflama oranı %83,5 olarak bulunmuştur. Toprak kayması şüphesi durumları beş düzeye ayrılmış ve bu düzeylerin oranları tespit edilerek haritada ilgili bölgeler farklı renk ile gösterilerek harita elde edilmiştir (76).

Lojistik regresyonun en yoğun kullanıldığı alanlardan birisi de sağlıktır. Özellikle “hasta-sağlam”, “ölü-sağ”, “düşük-orta-yüksek” gibi kategorilere gözlemlerin atanması sorunu sağlık alanında sürekli kullanılmaktadır. Çünkü bir olayın gerçekleşme olasılığı ve bir bağımsız değişkenin olayın gerçekleşme olasılığı üzerindeki etkisi birçok araştırmada cevaplanması gereken sorulardır. Lojistik regresyon analizinin varsayım gerektirmemesi, yorumlamanın kolay olması ve faktörlerin etkileri hakkında bilgi vermesi nedeniyle sıklıkla tercih edilmektedir. Ancak, sağlık alanında yapılan araştırmalar incelendiğinde yapılan analizlerde kurgulanan model ile ilgili detaylı bilgi verilmediği gözlenmektedir. Yalnızca modelde elde edilen odds değerleri bilgisi paylaşılmakta, ancak modelin anlamlılığı ve uyum iyiliği sonuçları ile ilgili yeterli bilgi sunulmamaktadır.

Tıp verileri ile yapılan çalışmalardan birisi bebeğin doğum ağırlığının tahmin edildiği ve çok sayıda bağımsız değişkenin olduğu bir anket verisidir. Toplam 128 gebe kadına ait verilerde anne ve babaya ait çeşitli ölçümler, gestasyonel bilgiler, demografik özellikler ve çeşitli biyokimya ölçümleri ile 16 adet bağımsız değişken kullanılmıştır. Bağımsız değişkenlerin katsayılarının anlamlılığı Wald testi ile yapılır. Doğrusal regresyon modellerinde t testi ile yapılan anlamlılık testine benzerdir. Analiz sonucunda modelin olabilirlik oran testi sonucu anlamlı ( $X^2=4,034$ ,  $p<0,025$ ) ve 0,05 yanılma düzeyinde Hosmer-Lemeshow test istatistiği  $X^2=1,424$  bulunarak modelin uyumunun iyi düzeyde olduğu bildirilmiştir (77).

Bir üniversitede öğrencilerin sigara içme alışkanlığını etkileyen bağımsız değişkenlerin tespit edilmesi amacıyla yapılan bir çalışmada toplam 13 adet değişken

kullanılmıştır. Sigara kullananların kullanmayanlara göre hangi durumlardan etkilendikleri belirlenmeye çalışılan çalışma toplam 600 öğrenci ile gerçekleştirilmiştir. Çalışmada sınıflandırma başarısı toplamda %77 ve modelin anlamlılığı  $X^2=124,05$  ( $p<0,001$ ) olarak hesaplanmıştır. Sigara içme nedenleri arasındaki en yüksek odds oranına sahip değişkenin “sigara içmenin statü kazandırdığı inancı” olarak bulunmuştur (78).

Farklı bir çalışmada ise üniversite öğrencilerinde alkol kullanımını etkileyen faktörlerin belirlenmesi amaçlanmıştır. Yaş aralığı 18 ile 35 olan ve bir özel üniversitede öğrenim gören toplam 150 öğrenci ile çalışma tamamlanmıştır. Aile bilgileri, gelir düzeyi, alkol ve sigara tüketimi ve sportif yaşam tarzı bilgilerini içeren toplam 16 bağımsız değişken modele dahil edilmiştir. Analiz sonucunda sınıflama tablosundaki ayırma başarısı oranı %87,3 olarak hesaplanmıştır. Modelin anlamlılık ve uyum düzeyinin yüksek olduğu bildirilmiştir. Etkin faktörler içerisinde en yüksek odds oranı değerinin “alkol kullanımının statü göstergesi olması” ve “alkol kullanımının dertlerden uzaklaştırması” düşüncelerine ait görülmüştür (79).

Kas-iskelet sistemindeki ağrıya neden olan faktörlerin belirlenmesi için yapılan bir fizyoterapi çalışmasında ülkenin farklı bölgelerinde rastgele seçilen 2852 olgu çalışmaya dahil edilmiştir. İki kategorili lojistik regresyon olarak kurgulanan modelde 15 adet bağımsız değişken kullanılmıştır. Modelin anlamlılığı yüksek ( $X^2=3101,74$ ,  $p<0,001$ ) ve uyumu oldukça iyi (Nagelkerke  $R^2=0,182$ ) bulunmuştur. Sınıflama tablosuna göre başarı yüzdesi %70,5 olarak hesaplanmıştır. Değişkenlerden sekiz tanesi 0,05 yanılma düzeyinde anlamlı bulunmuştur (80).

Tıp verileri kullanılan ve lojistik regresyon uygulaması detaylı olarak anlatılan başka bir çalışmada ise koroner arter hastalığına ait değişkenler kullanılmıştır. Toplam 500 hastadan alınan bilgiler ile 34 bağımsız değişken içeren bir model oluşturulmuştur. Demografik özellikler, EKG bulguları ve çeşitli biyokimya bilgileri ile analiz yapılmış ve model anlamlı bulunmuştur. Pearson  $X^2=140,65$  ve sapma istatistiği  $D=83,86$  bulunmuş olup anlamlıdır. Sınıflama tablosuna göre toplamda ayırma başarı oranı %90,0 şeklinde oldukça yüksek olarak bulunmuştur (81).

Sağlık alanında yalnızca tıp alanında lojistik regresyon analizi kullanılmamakta, diş hekimliği alanında da kullanılmaktadır. Buna örnek olarak bir

Diş Hekimliği Fakültesi Protetik Diş Tedavisi ABD’de tedavi edilen hastalar ile yapılan çalışma örnek verilebilir. Çalışmada toplam 350 hastaya ait bilgiler kullanılmış ve protez kullanımı üzerinde etkili olan faktörler araştırılmaya çalışılmıştır. Bağımsız değişkenler olarak demografik özellikler, alkol/sigara kullanımı, ağrı, gingiva durumu, kalp durumu, ağızdan soluma, ağrı, çürük, fırçalama alışkanlığı gibi toplam 17 değişken kullanılmıştır. Oluşturulan ilk modelde uyum ve anlamlılık düzeyi yüksek olmasına karşın sınıflandırma başarı oranı %57,7 gibi düşük değer hesaplanmıştır. Belirlenen sınıflar ile ilişkili olmayan değişkenler modelden çıkarıldığında daha anlamlı bir model bulunmuş ( $X^2=238,505$ ) ve adımsal yöntem ile ileriye dönük eleme seçeneği ile %83 oranında sınıflandırma başarısı elde edilmiştir. Çalışmada iki farklı örnek durum ile protez kullanma riski için tahmin sonuçları hakkında bilgi verilmiştir (82).

Lojistik regresyon uygulama çalışmaları esnasında zaman zaman çoklu bağlantı sorunu ortaya çıkabilmektedir. Bu durumda yapılan tahminlerin güven düzeyleri artar ve güvenilir olmayan tahminler yapılmasına neden olur. Bu sorunu gidermek için çeşitli yöntemler vardır. Bunlardan birisi de Hoerl & Kennard tarafından önerilen Ridge regresyonudur. Bu yöntemde, parametre tahmini yapılan çözüm matrisine küçük bir değer eklenerek çoklu bağlantı sorunu giderilmeye çalışılır. Örnek veri kümesi olarak böbrek nakli yapılan toplam 882 hastaya ait bilgiler kullanılmıştır. Bağımlı değişken olarak bu hastalarda huzursuz bacak sendromu oluşma olasılığı, bağımsız değişkenler olarak demografik özellikler, klinik özellikler ve çeşitli kan parametreleri alınmıştır. Değişkenler içerisinde aralarında korelasyonu yüksek olan ölçümler olduğundan modelde Ridge regresyon yöntemi kullanılmıştır. Bu tür yöntemler uygulandığında Hata Kareler Ortalaması (HKO) değeri küçük olan model tercih edilir. Çalışmada n:100 – n:600 arasında farklı örneklem büyüklükleri için modeller oluşturulmuş ve her bir model için hesaplanan HKO değerleri karşılaştırılmıştır. Örneklem büyüklüğü 300 ve az olan modellerde Ridge regresyon yönteminin daha kullanışlı olduğu bildirilmiştir (83).

Bu çalışmada iki farklı veri kümesi ile iki kategorili ve çok kategorili lojistik regresyon modelleri kurulmuştur. İki kategorili model için kalp krizi riski üzerinde etkili olan faktörler belirlenmeye çalışılmıştır. Modelde kalp krizi riski bağımlı değişken, demografik ve klinik bilgilerden oluşan toplam 14 bağımsız değişken

bulunmaktadır. Modelde adımsal ileriye dönük eleme yöntemi uygulanarak en uygun model elde edilmiştir. Omnibus  $X^2=200,99$  ve  $-2LL=216,64$  olarak hesaplanmıştır. Bu değerler modelin anlamlılığının oldukça yüksek olduğunu göstermektedir. Ayrıca,  $R^2$  değerleri de oldukça yüksek bulunmuştur. Nagelkerke  $R^2=0,648$  olarak hesaplanmış olup modelin uyum iyiliğini gösteren önemli göstergelerdendir. Modelin uyum iyiliği göstergelerinden birisi olan Hosmer-Lemeshow test sonucu  $X^2=10,54$  ve  $p=0,229$  olarak hesaplanmıştır. İleriye dönük elemelerde toplam 7 aşama elde edilmiştir. Son aşamaya ait sınıflama tablosu ayırma başarı oranı %84,3 olarak hesaplanmıştır. Gözlemlerin %84'ünün doğru sınıflara yerleştirilmiş olması modelin oldukça başarılı olarak kurulmuş olduğunu göstermektedir. Adımsal yöntem ile modeldeki çoklu bağlantı problemi de giderilmiştir. Böylece, parametre tahminlerinin hesaplanmasında Ridge regresyon veya benzer yöntem kullanılması gerek kalmamıştır. Yukarıda verilen örneklerde elde edilen model anlamlılık ve uyum sonuçları ile bu çalışmanın sonuçları benzer bulunmuştur. Ayrıca diğer çalışmalarda yapılan analizlerin benzerleri uygulanmış ve elde edilen tabloların yorumlanması benzer tarzda yapılmaktadır. Bu çalışmada, diğer çalışmalardan farklı olarak artık incelenmesi gerçekleştirilmiştir. Artıkların incelenmesi ile etkin, uzak ve aykırı gözlemlerin tespit edilmesi mümkün olmaktadır. Cook uzaklığı, kaldıraç noktaları,  $dfBeta$  veya  $dFits$  değerleri bu tespitlerin yapılmasında artıkların yanı sıra değerlendirilen ölçütlerdendir. Bir ölçüt için sorunlu bulunan yani uç değerlerin dışında olan değerler muhtemelen diğer ölçütlerde de uç değerlerin dışarısında olmaktadır. Lojistik regresyon analizi sonucunda lojit artıklar ile standardize artıklar en sık kullanılan artık türlerindendir. Lojit artıkların yanı sıra Cook uzaklığı ve kaldıraç noktaları ölçüt sınırları dışında olan değerleri içermektedir. Bu nedenle, bu değerlere ait gözlemler tespit edilmiştir. Öncelikle ölçüm sonuçlarının kontrolü yapılmış ve bu gözlemlerin sayısı 8 olup veri kümesinden çıkarılmasının örneklem sayısını etkileyemeyeceği düşünülmüştür. Etkin veya aykırı gözlemlerin çıkarılması ile model daha uyumlu ve anlamlı hale getirilmiştir. Omnibus  $X^2=248,52$  ve  $-2LL=157,96$  değerleri elde edilmiştir. Bu durumda, sınıflama başarı oranı %86,8'e yükselmiştir.

Çok kategorili model için sağlık sigortası ödemesi yapan müşterilerin demografik özellikleri kullanılarak yaşadıkları bölge tahmin edilmeye çalışılmıştır. Modelin anlamlılık değeri  $X^2=142,35$  ve  $-2LL=3560,95$  şeklinde hesaplanmıştır.

Ancak en çok olabilirlik değerin yüksek olması modeldeki gözlem sayısının fazla olmasından kaynaklanmaktadır. Ayrıca model uyumunu gösteren  $R^2$  değeri düşük bulunmuştur. Modelin sınıflama başarı oranı %32,7 olarak bulunmuştur. Fakat çok kategorili model oluşturulduğu düşünüldüğünde bu başarı oranının orta düzeyde olduğu görülmektedir.



## SONUÇ ve ÖNERİLER

Genel olarak bağımlı değişkenin sonucunu tahmin edebilecek en sade modeli bulmak için kullanılan lojistik regresyon analizi birçok alanda kullanılmaktadır. Kullanılmak için çok fazla varsayımı olmayan bağımlı değişkenin sürekli ve ya kategorik, iki ve ya ikiden fazla kategoride olmasının sorun teşkil etmediği lojistik regresyon analizi bu tez çalışmasında incelenmiştir.

İlk olarak bağımlı değişkenin iki kategorili olduğu Cleveland Clinic verisi olarak bilinen ve araştırmacılar tarafında sıklıkla tercih edilen kalp krizi riski veri kümesi kullanılmıştır.

Bağımlı değişkenin kategorik (yüksek MI riski, düşük MI riski), bağımsız değişkenlerin kategorik ve sürekli verilerden oluşan veri setine, tanımlayıcı istatistikler, bağımsız grup student t-testi, parametrik olmayan Mann Whitney U testi, tek yönlü varyans analizi (one-way ANOVA), Tukey HSD post-hoc düzeltmesi, ki-kare ve lojistik regresyon analizi uygulanmıştır.

Yapılan bu analiz sonuçlarına göre hastaların yaşı, dinlenme anı kan basıncı, maksimum nabız sayısı, ST segmenti çökme değeri ve görüntülenen damar sayısı, düşük ve yüksek MI risk sınıfları arasında anlamlı farklılık göstermiştir ( $p<0,001$ ).

Erkek hastaların oranı yüksek MI risk grubunda anlamlı düzeyde yüksek olup OR değeri 1,669 olarak hesaplanmıştır.

AKŞ oranları risk grupları arasında anlamlı bulunmamıştır. Dinlenme anı EKG bulgularında ST-T dalga anormalliği yüksek risk grubunda bulunan hastalarda daha yüksek oranda gözlenmiştir. Egzersiz anında anjina görülme oranı düşük risk grubundaki hastalarda daha yüksek oranda izlenmiştir.

Tam doymuş modelde genel tahmin başarısı %85,5 olarak hesaplanmıştır. Bu yüzde yüksek olmasına rağmen bulunan koşul indeks değerinin 50,785 değerine kadar artış göstermesi sebebi ile çoklu bağlantı probleminin olduğunu tespit edilmiştir.

Bunun için modeldeki değişken sayısını azaltacak model seçimi ile bu sorunun giderilmesine karar verilmiştir.

Olabilirlik oran yöntemine uygun ileriye dönük model seçimi (forward LR) tercih edilmiştir.

Erkek cinsiyet, egzersiz dönemi anjin, sabit ve tersine çevrilebilir ST defekti değişkenleri pozitif etkiye sahip bulunmuştur. Tipik göğüs ağrısı, ST çökmesi, SAT düz eğim ve damar sayısı katsayıları negatif etkiye sahip bulunmuştur.

Yeniden düzenlenen modelin Omnibus ki-kare değeri 248,52 ( $p<0,001$ ) ve model uyum değerleri  $-2LL=157,96$ ,  $R^2_{Cox\&Snell}=0,569$  ve  $R^2_{Nagelkerke}=0,761$  bulunmuştur. Hosmer& Lemeshow ki-kare=10,348 ve  $p=0,241$  olduğundan model anlamlılığı mevcuttur.

Etkin gözlemlerin modelden çıkarılmasıyla atipik göğüs ağrısı kategorisinin de modele anlamlı katkı yaptığı görülmüştür. Erkek cinsiyet ve egzersiz esnasında mevcut olan göğüs ağrısı modele pozitif katkı yaparken diğer anlamlı değişkenlerin tamamı negatif katkı yapmaktadır.

Çalışmanın ikinci örneğinde ise bağımlı değişkenin dört kategorili (Güneydoğu, Güneybatı, Kuzeydoğu, Kuzeybatı) olduğu sağlık sigortası harcaması veri kümesi kullanılmıştır. Bölgelere göre yaş, cinsiyet, vki, sigortadan faydalanan çocuk sayısı, sigara kullanımı, sağlık sigortası harcaması bağımsız değişkenleri karşılaştırılmıştır. Ülkenin dört farklı bölgesinde yaşayan sigorta müşterilerinin özelliklerinin etkisini belirlemek amacıyla çok kategorili (multinomial) lojistik regresyon analizinin yanı sıra tanımlayıcı istatistikler, tek yönlü varyans analizi, parametrik olmayan Kruskal-Wallis H testi yapılmıştır.

Çıkan analiz sonuçlarına göre müşterilerin yaşları ve çocuk sayıları bölgeler arasında anlamlı farklılığa sahip bulunmamıştır ( $p>0,05$ ).

Müşterilerin VKİ değerleri güneydoğu bölgesinde yaşayanlar için anlamlı düzeyde yüksektir ( $p<0,001$ ).

Müşterilerin ödediği sağlık sigorta miktarı da güneydoğu bölgesi müşterilerinde anlamlı düzeyde yüksek bulunmuştur ( $p=0,010$ ).

Yalnızca sabit katsayının olduğu modelde modelin AIC ve BIC değerleri sırasıyla 3602,95 ve 3712,13'e gerilemiştir. Buna göre  $-2LL$  değeri 3560,95

hesaplanmıştır. Modelin anlamlılık ki-kare istatistiği 145,35 ve  $p<0,001$  bulunduğundan oluşturulan model anlamlıdır.

Modelin uyum düzeyinin anlamlı olmadığı görülmüştür. Çünkü hem Pearson hem de Sapma ki-kare istatistikleri anlamlı bulunmamış ve sıfır hipotezi reddedilememiştir ( $p>0,05$ ).

Bağımsız değişkenler eklenince modelin uyumunun ve anlamlılığının hangi düzeyde olduğunu belirlemek için Olabilirlik Oran Testi sonuçları incelenmiş müşterilere ait VKİ, sigorta ücreti ve sigara kullanımı ki-kare istatistiklerinin anlamlı bulunması nedeniyle bu değişkenlerin modelde anlamlı olabileceği düşünülmüştür.

VKİ değerleri Güneydoğu bölgesinde yaşıyor olma olasılığını Kuzeybatı bölgesinde yaşayan müşterilere göre 1,13 kat artırmaktadır.

Çocuk sayısı, sigorta ücreti ve cinsiyetin Güneydoğu bölgesinde yaşama olasılığı üzerinde anlamlı etkisi bulunmamıştır.

Sigara kullanmıyor olma durumunun Güneydoğu bölgesinde yaşıyor olma olasılığını azaltıcı etkisi bulunmuştur.

En iyi tahmin oranları Güneydoğu bölgesine ait olanlardır (%59,1). Diğer bölgeler için yapılan tahminlerin başarı oranları oldukça düşük ve genel olarak modelin genel başarı oranı %32,7 bulunmuştur.

Artıkların incelenmesi gerçekleştirilmiştir ancak çıkarılması gereken gözlem değerleri bulunmamıştır.

Gerek uzaklıklar gerekse artıklardan aşırı büyük veya aşırı küçük değerlere rastlanmamıştır. Bu nedenle, gözlemlerin çıkarılması ile anlamlılık düzeyi daha yüksek bir model elde edilememiştir. Ayrıca, VIF ve koşul indeksi değerlerinin incelenmesine göre modelde çoklu bağlantı problemi de bulunmamaktadır.

Çalışmada kullanılacak veriler her zaman sayısal veriler olmayıp kategorik verilerden de oluşabilmektedir. Bağımlı değişken ister iki kategorili ister çok kategorili olsun, ister nümerik ister kategorik yapıda olsun bu gibi durumlarda LRA araştırmacılar için çok kullanışlı bir analiz haline gelmekte ve kolay yorumlanabilme açısından bu analizin kullanımı önerilmektedir.

## KAYNAKLAR

1. Gürcan M. Lojistik regresyon analizi ve bir uygulama. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Samsun, (Tez Danışmanı: Yrd. Doç. Dr. Yüksel Öner), 1998.
2. Finney DJ. Probit analysis: a statistical treatment of the sigmoid response curve. 7th ed. Cambridge University Press, Cambridge, 1972, p. 33.
3. Hüdaverdi B. Lojistik regresyon analizi: tıp verileri üzerine bir uygulama. Kocaeli Üniversitesi Sosyal Bilimler Dergisi 2004; 2: 185–208.
4. Kaya M, Genç M, Kaya B, Pehlivan E. Tıp fakültesi ve sağlık yüksekokulu öğrencilerinde depresif belirti yaygınlığı, stresle başa çıkma tarzları ve etkileyen faktörler. Türk Psikiyatri Dergisi 2007; 18(2): 137–46.
5. Girginer N, Cankuş B. Tramvay yolcu memnuniyetinin lojistik regresyon analiziyle ölçülmesi: estram örneği. Yönetim ve Ekonomi Araştırmaları Dergisi 2008; 5(1).
6. Çakıcı İ, Oğuzhan A, Özdil T. Temel İstatistik 1. 3. Baskı, İstanbul: Özal Matbaası, 2000.
7. Özarıcı Ö. Farklı not sistemlerinde öğrencinin başarılı olma olasılığının probit regresyon analiziyle değerlendirilmesi. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Eskişehir, (Tez Danışmanı: Prof. Dr. Necla Çömlekçi), 1996; 6.
8. Özdamar K. Spss ile Biyoistatistik. 5. Baskı. Eskişehir: Kaan Kitabevi, 2003.
9. Vural A. Aykırı değerlerin regresyon modellerine etkileri ve sağlam kestiriciler. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul, (Tez Danışman: Prof. Dr. Selahattin Güriş), 2007; 2.
10. Gökçe A. Bazı sağlam tahmincilerle dayalı yanlı regresyon analizi. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Eskişehir,

(Tez Danışmanı: Prof. Dr. Özlem Alpu), 2019.

11. Kılıç S. Lojistik regresyon analizi ve pazarlama araştırmalarında bir uygulama. , İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, (Tez Danışmanı: Doç. Dr. Burç Ülengin), 2000.
12. Özdamar K. Paket Programlar ile İstatistiksel Veri Analizi. Ankara: Sözkese Matbaacılık, 2015, s. 541–543.
13. Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. Ankara: Detay Yayıncılık; 2013, s. 637–710.
14. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley; 2000.
15. Seven Z. Değişken seçimi yöntemi olarak adımsal lojistik regresyon ile adımsal diskriminant analizinin karşılaştırılması. Gazi Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara, 1997.
16. Hosmer DW, Lemeshow S. A Goodness of Fit Tests for the Multiple Logistic Regression Model. A10, Communications in Statistics, 1980, p. 1043–1069.
17. Cook D, Dixon P, Duckworth WM, Kaiser MS, Koehler K, Meeker WQ, Stephenson WR. Binary response and logistic regression analysis. 2001; 1–23.
18. Field A. Discovering Statistics Using SPSS. 2nd Ed., 2005.
19. <https://covid19.saglik.gov.tr/TR-66475/insidans.html>.
20. Kuyucu YE. Lojistik regresyon analizi (lra), yapay sinir ağları (ysa) ve sınıflandırma ve regresyon ağaçları (c&rt) yöntemlerinin karşılaştırılması ve tıp alanında bir uygulama. Gaziosmanpaşa Üniversitesi Sağlık Bilimleri Enstitüsü, Yüksek Lisans tezi, Tokat, (Tez Danışmanı: Yrd. Doç. Dr. Ünal Erkorkmaz), 2012.
21. Kaşko Y. Çoklu bağlantı durumunda ikili (binary) lojistik regresyon modelinde gerçekleşen 1. tip hata ve testin gücü. Ankara Üniversitesi Fen Bilimleri

- Enstitüsü, Yüksek Lisans Tezi, Ankara, (Tez Danışmanı: M. Muhip Özkan), 2007.
22. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied Linear Statistical Models. New York, America: McGraw-Hill, 2005, p. 563–593.
  23. Agresti A, Finlay B. Statistical Methods for the Social Sciences. Dellen, 1986, p. 576–580.
  24. Elhan AH. Lojistik regresyon analizin incelenmesi ve tıpta bir uygulaması. Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara, (Tez Danışmanı: Prof. Dr. Yıldır Atakurt), 1997.
  25. Ayhan S. Sıralı lojistik regresyon analiziyle Türkiye’deki hemşirelerin iş bırakma niyetini etkileyen faktörlerin belirlenmesi. Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Eskişehir, (Tez Danışmanı: Yrd. Doç. Dr. Zeki Yıldız), 2006.
  26. Kleinbaum DG, Klein M. Logistic Regression: A Self-Learning Text. 2nd Ed., New York; 2002.
  27. Arabacı Ö. Lojistik regresyon analizi ve bir uygulama denemesi. Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Bursa, (Tez Danışmanı: Prof. Dr. Mustafa Aytaç), 2002.
  28. Şensoy ZE. Nonlineer lojistik regresyon ve uygulaması. Marmara Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, (Tez Danışmanları: Prof. Dr. Aydın Erar, Prof. Dr. Müjgan Tez), 2009.
  29. Kara ÖS. Lojistik regresyon analizi ve kadın işgücü üzerine bir uygulama. Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Bursa, (Tez Danışmanı: Prof. Dr. Mustafa Aytaç), 2015.
  30. Tatlıdil H. Uygulamalı Çok Değişkenli İstatistiksel Analiz. Ankara: Cem Ofset, 1996.

31. Hu S. Modelling trip generation / trip accessibility using logit models. Edinburgh Napier University, Doctoral Thesis. 2010.
32. Weisberg S. Applied Linear Regression. 4th Ed., Hoboken, New Jersey, ABD: John Wiley & Sons, 2014.
33. Sharma S. Applied Multivariate Techniques. New York: John Wiley and Sons Inc, 1996.
34. Christensen R. Log-Linear Models and Logistic Regression. New York: Springer-Verlag Inc, 1997.
35. Box GEP, Tidwell PW. Transformation of the independent variables. Technometrics 1962; 4(4): 531–550.
36. Albayrak A. Uygulamalı Çok Değişkenli İstatistik Teknikleri. Ankara: Asil Yayın Dağıtım, 2006.
37. Kalaycı Ş. SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri. Ankara: Asil Yayın Dağıtım, 2010.
38. Montgomery DC, Peck EA, Vining GG. Doğrusal Regresyon Analizine Giriş. Ankara: Nobel Akademik Yayıncılık, 2013.
39. Gamgam H, Altunkaynak B. SPSS Uygulamalı Regresyon Analizi. Seçkin Yayıncılık; 2015.
40. Wilson JR, Lorenz KA. Modeling Binary Correlated Responses Using SAS, SPSS and R, ICSA Book Series In Statistics 9. Switzerland: Springer Publishing, 2015.
41. Aydın N, Arı E. Hanchalkı otomobil talebini belirleyen etkenlerin ikili lojistik regresyon yöntemiyle analizi: Türkiye örneği. Kastamonu Üniversitesi, İİBF Dergisi 2016; 12(1): 1–22.
42. Tanju Ö. Cluster based model diagnostic for logistic regression. Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara, (Tez Danışmanı: Doç. Dr.

Zeynep Işıl Kalaylıoğlu Akyıldız), 2016.

43. Field A. Discovering Statistics Using SPSS. London: Sage, 2009.
44. Polat H. Lojistik regresyon analizi: sağlık verileri üzerine bir uygulama. Recep Tayyip Erdoğan Üniversitesi Sosyal Bilimler Enstitüsü, Rize, (Tez Danışmanı: Prof. Dr. Ali Sait Albayrak), 2017.
45. Önder H, Cebeci Z. Lojistik regresyonlarda değişken seçimi. Çukurova Üniversitesi Ziraat Fakültesi Dergisi 2002; 17(2): 105–114.
46. Ergün M. Bilimsel Araştırmalarda Bilgisayarla İstatistik Uygulamaları. Ankara: Ocak Yayınları, 1995.
47. Baydemir MB. Lojistik regresyon analizi üzerine bir inceleme. İnönü Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri Ana Bilim Dalı, Yüksek Lisans Tezi, Malatya, (Tez Danışmanı: Prof. Dr. Mehmet Güngör), 2014.
48. Öztürk S. Yeni doğan bebeklerde düşük doğum ağırlığının ikili lojistik regresyonla en çok olabilirlik ve mars yaklaşımına dayalı modellenmesi. Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Muğla, (Tez Danışmanı: Yrd. Doç. Dr. Volkan Sevinç), 2012.
49. Kleinbaum DG, Klein M. Logistic Regression: A Self-Learning Text. New York: Springer-Verlag, 2010, p. 304.
50. Cöhce MO. Lojistik regresyon analizi ile üniversite öğrencilerinin kariyer planlarını etkileyen faktörlerin analizi. Bursa Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri Ana Bilim Dalı, Bursa, (Tez Danışmanı: Prof. Dr. Nuran Bayram Namlı), 2020.
51. Hilbe JM. Practical Guide to Logistic Regression. New York: Taylor&Francis Group, 2016, p. 88.
52. Çokluk Ö. Lojistik regresyon analizi: kavram ve uygulama. Kuram ve Uygulamada Eğitim Bilimleri 2010; 10(3): 1357–1407.

53. Collet D. Modelling Binary Data. 2nd Ed., New York: Chapman & Hall; 2003.
54. Ahmad S, Midi H, Ramli NM. Diagnostics for residual outliers using deviance component in binary logistic regression. World Applied Sciences Journal 2011; 14(8): 1125–1130.
55. Chatterjee S, Hadi AS. Regression Analysis by Example. 5th Ed., New York: John Wiley & Sons, 2012.
56. Chatterjee S, Hadi AS. Regression Analysis by Example. New York: Wiley-Interscience, 2006.
57. Montgomery CD, Peck AE, Vining G. Introduction to Linear Regression Analysis. New York: John Wiley & Sons; 2001.
58. Acarlar I. Varyansın sağlam tahmin edicilerine dayalı cook uzaklığı istatistiği'nin incelenmesi. Süleyman Demirel Üniversitesi Fen Bilim Enstitüsü Dergisi 2011; 15(1): 8–14.
59. Türker H. Lojistik regresyon ve uygulamaları. Çanakkale Onsekiz Mart Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Çanakkale, (Tez Danışmanı: Bilgehan Güven), 2016.
60. Akyol K, Şen B, Çalık E. Biyokimya ve hemogram laboratuvar test sonuçlarının lojistik regresyon yöntemiyle analiz. Akad Bilişim Konferansı 2012; 345–355.
61. Bozkurt B. Kredi ve yurtlar kurumunda kalan öğrencilerin memnuniyet derecelerinin lojistik regresyon yöntemi ile araştırılması: Edirne ili örneği. Trakya Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Edirne, (Tez Danışmanı: Prof. Dr. Ayşe Akyol), 2011.
62. Akçay A. Broiler Coccidiosis'inde risk faktörlerinin lojistik regresyon analizi ile belirlenmesi. Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Doktora Tezi, Ankara, (Tez Danışmanı: Prof. Dr. Okan Ertuğrul), 2009.
63. Liao TF. Interpreting Probability Models: Logit, Probit and Other Generalized

Linear Models. CA: Sage, 1994, p. 88.

64. Begg CB, Gray R. Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika* 1984; 71(1): 11–13.
65. Arı E. Sıralı lojistik regresyonda paralel doğrular varsayımı ve çözümleme yaklaşımları. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Eskişehir, (Tez Danışmanı: Prof. Dr. Zeki Yıldız), 2013.
66. Long JS. *Regression Models for Categorical and Limited Dependent Variables*. CA: Sage, 1997.
67. Scott SC, Goldberg MS, Mayo NE. Statistical Assessment of ordinal outcomes in comparative studies. *Clinical Epidemiology* 1997; 50(1): 45–55.
68. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd Ed. London: Chapman & Hall, 1989.
69. Bender R, Benner A. Calculating ordinal regression models in SAS and S-Plus. *Biometrical Journal* 2000; 42(6): 677–699.
70. İnal M, Topuz D, Uçan O. Doğrusal olasılık ve logit modelleri ile parametre tahmini. *Sosyoekonomi* 2006; 3(3): 48–72.
71. Makalic E, Schmidt DF. Review of modern logistic regression methods with applicagtion to small and medium sample size problems. In: 3rd Australian Joint Conference Advances in Artificial Intelligence Proceedings. Berlin, Germany; 2010.
72. Patnaik D, Sharma NS. Selection problems for application of Probit, Tobit, Logit & Maximum Likelihood estimation: a methodological issue. *International Journal of Electrical and Computer Engineering* 2013; 3(7): 13–29.
73. Güneri O, Durmuş B. Dependent dummy variable models: an application of Logit, Probit and Tobit models on survey data. *International*

Journal of Computational and Experimental Science and Engineering 2020; 6(1): 63–74.

74. Aktaş C, Erkuş O. Lojistik regresyon analizi ile Eskişehir'in sis kestiriminin incelenmesi. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi 2009; 8(16): 47–59.
75. Ohlmacher GC, Davis JC. Using multiple logistic regression and gis technology to predict lanslide hazard in northeast kansas, USA. Engineering Geology 2003; 69: 331–343.
76. Ayalew L, Yamagishi H. The application of gis-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains. Geomorphology 2005; 65: 15–31.
77. Bircan H. Lojistik regresyon analizi: tıp verileri üzerine bir uygulama. Kocaeli Üniversitesi Sosyal Bilimler Dergisi 2004; 2: 185–208.
78. Aktaş C. Lojistik regresyon analizi: öğrencilerin sigara içme alışkanlığı üzerine bir uygulama. Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 2009; 1(26): 107–122.
79. Ersöz T, Kapanşahin G, Ersöz F. Investigation of the alcohol usage habits of university students by logistic regression analysis. Karabük Üniversitesi Sosyal Bilimler Enstitü Dergisi 2019; 7(1): 42–50.
80. Cımbız A, Uzgören N, Aras Ö, Öztürk S, Elem E, Aksoy CC. Kas İskelet sisteminde ağrıya ait risk faktörlerinin lojistik regresyon analizi ile belirlenmesi: pilot çalışma. Türkiye Fiziksel Tıp ve Rehabilitasyon Dergisi 2007; 18(1): 20–27.
81. Atakurt Y. Lojistik Regresyon analizi ve tıp alanında kullanımına ilişkin bir uygulama. Ankara Üniversitesi Tıp Fakültesi Mecmuası 1999; 52(4): 191–199.
82. Coşkun S, Kartal M, Bircan H. Lojistik regresyon analizinin incelenmesi ve diş hekimliğinde bir uygulama. Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi

Dergisi 2004; 7(1): 42–50.

83. Vago E, Kemeny S. Logistic ridge regression for clinical data analysis (a case study). *Applied Ecology And Environmental Research* 2006; 4(2): 171–179.



## ÖZGEÇMİŞ

### **Kişisel Bilgiler**

Ad / Soyad : Büşra ÇATAL  
Doğum Yeri :  
Doğum Tarihi :  
Telefon No :  
E- mail :

### **Eğitim Bilgileri**

Lise : Gül - Çetin Kaur Lisesi  
Lisans : Akdeniz Üniversitesi Hemşirelik Yüksekokulu (2011 - 2015)  
Yüksek Lisans : Süleyman Demirel Üniversitesi Tıp Fakültesi, Biyoistatistik ABD (2019-2021)

### **İş Deneyim Bilgileri**

Antalya Medstar Hastanesi, Hemşire, Kemik İliği Transplantasyonu, 2015  
İstanbul Okmeydanı Eğitim ve Araştırma Hastanesi, Hemşire, Ameliyathane, 2015 - 2017