



T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**BİNOM DAĞILIMLI YANIT DEĞİŞKENİ
İÇİN GENELLEŞTİRİLMİŞ LİNEER
MODELLER: KANSER VE COVID-19
PANDEMİ VERİLERİ ÜZERİNE BİR
UYGULAMASI**

Esra SARI

YÜKSEK LİSANS

İstatistik Anabilim Dalı

Kasım-2023
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Esra SARI tarafından hazırlanan “Binom Dağılımlı Yanıt Değişkeni İçin Genelleştirilmiş Lineer Modeller: Kanser ve COVID-19 Pandemi Verileri Üzerine Bir Uygulaması” adlı tez çalışması 13/11/2023 tarihinde aşağıdaki jüri tarafından oy birliği ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Dr.Öğr.Üyesi Ümran Münire KAHRAMAN

.....

Danışman

Doç. Dr. Neslihan İYİT

.....

Üye

Dr.Öğr.Üyesi Ümran Münire KAHRAMAN

.....

Üye

Dr.Öğr.Üyesi Aynur YONAR

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Ömer Faruk YÜKSEL
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

Esra SARI

Tarih:

ÖZET

YÜKSEK LİSANS

BİNOM DAĞILIMLI YANIT DEĞİŞKENİ İÇİN GENELLEŞTİRİLMİŞ LİNEER MODELLER: KANSER VE COVID-19 PANDEMİ VERİLERİ ÜZERİNE BİR UYGULAMASI

Esra SARI

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Doç.Dr. Neslihan İYİT

2023, 54 Sayfa

Jüri

Doç.Dr. Neslihan İYİT

Dr.Öğr.Üyesi Ümran Münire KAHRAMAN

Dr.Öğr.Üyesi Aynur YONAR

Bu çalışmada, küresel olarak 179 dünya ülkesinin COVID-19 pandemisinden kaynaklanan ölüm verileri ve kanser ile ilişkili olabilecek bütün tümör ve beraberindeki bulaşıcı olmayan hastalıklardan (BOH) kaynaklanan engelliliğe uyarlanmış yaşam yılları (disability adjusted life years) (DALY) verileri ile arasındaki ilişkiler genelleştirilmiş lineer model (GLM) yaklaşımı ile incelenmiştir.

Bütün dünyada çok sayıda insanı etkileyen bir salgın olan COVID-19 pandemisi, Dünya Sağlık Örgütü (DSÖ) tarafından 11 Mart 2020 tarihinde pandemi olarak ilan edilmiştir. DALY verileri, erken ölümler nedeniyle kaybedilen yaşam yılları ve engelli olarak yaşanan yıllardan kaynaklanan toplam hastalık yükünü ölçmektedir. Bu çalışmada, açıklayıcı değişkenler olarak kanser ile ilişkili olabilecek tümör ve beraberindeki BOH'lardan kaynaklanan DALY verileri kategorik hale getirilerek kullanılmıştır. COVID-19 pandemisinin neden olduğu toplam ölüm sayısı kategorik hale getirilmiş ve daha sonra yanıt değişkeni olarak çalışmaya alınmıştır. Bu çalışmada istatistiksel yöntemler olarak GLM ailesindeki ikili lojit (binary logit) ve ikili probit (binary probit) regresyon modelleri kullanılarak, kanser ile ilişkili olabilecek tümör ve eşlik eden BOH'ların COVID-19 pandemisi üzerindeki etkileri araştırılmıştır. Bu çalışmaya özel olarak, hataların istatistiksel anlamda normal bir dağılıma sahip olduğu varsayımına dayanan probit modelinin, hataların lojistik bir dağılıma sahip olduğu varsayımına dayanan lojit modeline üstünlüğü vurgulanmıştır. Bu çalışmanın nihai sonucu olarak, kanser ile ilişkili olabilecek bütün tümör ve beraberindeki bulaşıcı olmayan hastalıklardan (BOH) kaynaklanan DALY verilerininin COVID-19 pandemisinin neden olduğu toplam ölüm verileri üzerindeki etkisi istatistiksel olarak anlamlı bulunmuştur.

Anahtar Kelimeler: Bulaşıcı olmayan hastalıklar, COVID-19 pandemisi, genelleştirilmiş lineer modeller, hastalık yükü, lojit model, probit model.

ABSTRACT

MS THESIS

GENERALIZED LINEAR MODELS FOR BINOMIALLY DISTRIBUTED RESPONSE VARIABLE: AN APPLICATION ON CANCER AND COVID-19 PANDEMIC DATA

Esra SARI

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
SELÇUK UNIVERSITY
THE DEGREE OF MASTER OF PHILOSOPHY
IN STATISTICS**

**Advisor: Assoc.Prof.Dr. Neslihan İYİT
Year, 54 Pages**

**Jury
Assoc.Prof.Dr. Neslihan İYİT
Assist.Prof.Dr. Ümran Münire KAHRAMAN
Assist.Prof.Dr. Aynur YONAR**

In this study, the relationship between mortality data from the COVID-19 Pandemic of globally 179 countries in the world and DALY (disability adjusted life years) data from all tumors that may be associated with cancer, and accompanying non-communicable diseases (NCDs) is statistically analyzed using the generalized linear models (GLM) approach.

The COVID-19 pandemic, which is a disease epidemic that affects many people all over the world, is declared a pandemic by the World Health Organization (WHO) on March 11, 2020. DALY measures the total disease of burden resulting from years of life lost due to premature death and years of disability. In this study, DALY data from all tumors that may be associated with cancer, and accompanying NCDs are used as explanatory variables by categorizing them. The total number of deaths caused by the COVID-19 pandemic is made categorical and then taken as the response variable into the study. In this study, the effects of all tumors that may be associated with cancer, and accompanying NCDs on the COVID-19 pandemic are investigated by using binary logit and binary probit regression models in the GLM family as the popular statistical methods. Specifically for this study, the superiority of the probit model based on the assumption that the errors have normal distribution over the logit model based on the assumption that the errors have logistic distribution is emphasized. As a final conclusion of this study, the effect of DALY data from all tumors that may be associated with cancer, and accompanying non-communicable diseases (NCDs) on the total death data caused by the COVID-19 pandemic is found to be statistically significant.

Keywords: Non-communicable diseases, COVID-19 Pandemic, generalized linear models, burden of disease, logit model, probit model.

ÖNSÖZ

Tez çalışmamın her aşamasında bilgi, ilgi ve desteğini benden esirgemeyen çok değerli danışmanım Doç.Dr. Neslihan İYİT'e, tez savunma komitesi üyeleri Dr.Öğr.Üyesi Aynur YONAR'a ve Dr.Öğr.Üyesi Ümran Münire KAHRAMAN'a, bugünlere gelmemde en büyük katkıyı sağlayan ve her zaman yanımda olan sevgili aileme ve yakın dostum Ülkü ÖĞÜNMEZ'e sonsuz teşekkür ederim.

Esra SARI
KONYA-2023



İÇİNDEKİLER

| | |
|---|-------------|
| ÖZET | iv |
| ABSTRACT..... | v |
| ÖNSÖZ | vi |
| İÇİNDEKİLER | vii |
| SİMGELER VE KISALTMALAR | viii |
| 1. GİRİŞ | 1 |
| 2. KAYNAK ARAŞTIRMASI | 3 |
| 2.1. Genelleştirilmiş lineer modeller ile ilgili kaynak araştırması | 3 |
| 2.2. COVID-19 Pandemisi, bulaşıcı olmayan hastalıklar (BOH) ve engelliliğe uyarlanmış yaşam yılları (DALY) ile ilgili literatür taraması | 4 |
| 3. MATERYAL VE YÖNTEM..... | 6 |
| 3.1. Çok değişkenli lineer regresyon modeli | 6 |
| 3.2. Genelleştirilmiş lineer modeller | 7 |
| 3.2.1. Yanıt değişkenin dağılımı | 7 |
| 3.2.2. Sistemik bileşen | 8 |
| 3.2.3. Link (bağlantı) fonksiyonu | 8 |
| 3.3. Binom dağılımlı yanıt değişkeni için genelleştirilmiş lineer modeller | 10 |
| 3.3.1. Lojistik regresyon modeli | 10 |
| 3.3.2. Probit regresyon modeli | 11 |
| 3.4. Genelleştirilmiş lineer modellerde parametre tahmini | 12 |
| 3.4.1. Genelleştirilmiş lineer modellerde olabilirlik fonksiyonu | 13 |
| 3.4.2. Newton-Raphson metodu | 14 |
| 3.4.3. Fisher-Scoring metodu | 16 |
| 3.5. Genelleştirilmiş lineer modellerde bilgi kriterleri | 21 |
| 4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA..... | 23 |
| 4.1. Araştırma Sonuçları | 23 |
| 4.2. Tartışma | 32 |
| 5. SONUÇLAR VE ÖNERİLER | 39 |
| 5.1 Sonuçlar | 39 |
| 5.2 Öneriler | 41 |
| KAYNAKLAR | 43 |

SİMGELER VE KISALTMALAR

Kısaltmalar

Kısaltma: **Açıklama**

BOH: Bulaşıcı Olmayan Hastalık

DALY: Engelliliğe Uyarlanmış Yaşam Yılı (Disability Adjusted Life Year)

DSÖ: Dünya Sağlık Örgütü

GLM: Genelleştirilmiş Lineer Model (Generalized Linear Model)



1. GİRİŞ

Lineer regresyon modeli, bir yanıt değişkeni ile bir ya da birden fazla açıklayıcı değişken arasındaki istatistiksel ilişkileri araştırmak üzere “lineerlik”, “normallik” ve “varyansların homojenliği” varsayımları altında geleneksel olarak kullanılan istatistiksel modeldir (Weisberg, 2005; Freund ve ark., 2006; Su ve ark., 2012; Agresti, 2015; Fox, 2015; Montgomery ve ark., 2021).

Ancak gerçek dünya verilerinde modellemeye alınacak yanıt değişkeni her zaman normal dağılıma sahip değildir. Örneğin, kansere dair incelenecek verilerde tümörlerin iyi huylu ya da kötü huylu olması, belirli bir bölgedeki trafik kaza sayıları, bir şehirdeki bebek ölüm sayıları gibi verilerin normal dağılıma sahip olmayacağı aşikârdır (Rodríguez, 2008; Bourne ve ark., 2009; Geedipally ve ark., 2012; Sadigh ve ark., 2012; Zha ve ark., 2016; Lim-Dunham ve ark., 2019; İyit ve ark., 2021).

Nelder ve Wedderburn (1972) tarafından ortaya atılan genelleştirilmiş lineer model (GLM) kavramı lineer regresyon modelini de ihtiva eden bir model sınıfı olup, üstel dağılım ailesi sayesinde yanıt değişkeninin normal dağılımdan gelme varsayımını ortadan kaldırmıştır. Böylece GLM yaklaşımı ile binom dağılımı, Poisson dağılımı, gamma dağılımı gibi üstel dağılım ailesinin birer üyesi olan dağılımlara sahip yanıt değişkenleri için istatistiksel modelleme imkânı sağlanmıştır (McCullagh ve Nelder, 1989; İyit, 2018; Gill ve Torres, 2019).

11 Mart 2020’de Dünya Sağlık Örgütü (DSÖ) tarafından resmi olarak salgın ilan edilen COVID-19 pandemisi dünya çapında çok sayıda insanı etkilemiş, yüzbinlerce insanın ölümüne sebep olmuştur (Ciotti ve ark., 2020; Lone ve Ahmad, 2020; Tuttle, 2020; Maeda ve Nkengasong, 2021; Suryasa ve ark., 2021).

Engelliliğe ayarlanmış yaşam yılları (DALY), bir popülasyonda yaygın olarak görülen bir hastalık ya da sağlık durumu nedeniyle erken ölümlerden kaybedilen yaşam yılları ile engellilikle yaşanan yılların toplamı olup "hastalık yükünü" ölçen bir ölçüttür (Anand ve Hanson, 1997; WHO, 2019). DALY kavramı ilk olarak 1993 yılında “Dünya Kalkınma Raporu”nda (World Development Report) ortaya atılmıştır (World_bank, 1993; Chen ve ark., 2015). Bir birim DALY verisi, sağlıklı bir yılın kaybı olarak kabul edilir. DALY verileri ile çeşitli hastalık ve yaralanmalara bağlı erken ölümlerin neden olduğu hastalık yükü ile ölümlerle sonuçlanmayan ancak uzun süreli sakatlık ve fonksiyon kaybına neden olan hastalık durumları tek bir kriter aracılığı ile özetlenebilmektedir (Akman ve Civek, 2022).

Bulaşıcı olmayan hastalıklar (BOH), temelde akut bir enfeksiyonun neden olmadığı, tedavisi uzun zaman alan ve sağlık sorunlarına yol açabilen kanserler, kronik karaciğer hastalıkları, kardiyovasküler hastalıklar, diyabet vb. hastalıkların genel ifadesidir (Boutayeb ve Boutayeb, 2005; Wagner ve Brath, 2012; Horton, 2013; Islam ve ark., 2014).

Bu çalışmada uygulama olarak 179 dünya ülkesine ait COVID-19 pandemisi verileri ile kanser ve ilişkili olabilecek bütün tümör ve beraberindeki bulaşıcı olmayan hastalıklardan (BOH) kaynaklanan veriler arasındaki ilişkiler genelleştirilmiş lineer model (GLM) yaklaşımı ile incelenmiştir. Bu bağlamda yanıt değişkeni ile açıklayıcı değişkenler kategorik hale getirilerek, lojit ve probit regresyon modelleri oluşturulup, model sonuçları istatistiksel olarak karşılaştırılmıştır.

Bu çalışma beş kısımdan oluşmaktadır. Birinci kısım bu çalışmaya ait giriş kısmı olup ikinci kısımda GLM yaklaşımı, BOH ve DALY kavramları ile ilgili literatürde bulunan kaynak taramaları özet olarak verilmiştir. Üçüncü kısımda ise GLM'in genel yapısı tanıtılmış olup, tez çalışmasında kullanılan GLM'in özel bir hali olan "lojit" ve "probit" modeller ayrıntılı bir şekilde verilmiştir. Dördüncü kısım bu çalışmanın uygulama kısmı olup, yanıt değişkeni ve açıklayıcı değişkenler tanıtılmış, tanımlayıcı istatistikleri verilmiş ve bu değişkenler arasındaki istatistiksel ilişkileri incelemek üzere lojit ve probit regresyon modelleri oluşturulmuştur. Bu çalışmanın son kısmı olan beşinci kısımda ise modellerden elde edilen sonuçlar istatistiksel olarak karşılaştırılmış, anlamlı bulunan model sonuçları yorumlanmış ve elde edilen bu sonuçlar ışığında global olarak ele alınan dünya ülkeleri için önerilerde bulunulmuştur.

2. KAYNAK ARAŞTIRMASI

Bu çalışmada GLM, COVID-19 pandemisi ve BOH kaynaklı DALY kavramı ile ilgili literatür taramaları verilmiştir.

2.1. Genelleştirilmiş lineer modeller ile ilgili kaynak araştırması

GLM ilk olarak Nelder ve Wedderburn (1972) tarafından ortaya atılmıştır. Wedderburn (1974) GLIM (generalized linear interactive modeling) adında bir uygulama geliştirerek GLM yaklaşımı ile oluşturulabilecek karmaşık modellerin kolayca tahmin edilmesini sağlamıştır.

Probit regresyonun tarihçesi ise 1934 yılına dayanmaktadır. Bliss (1934) yanıt değişkeni iki değer alan (binary) modellerde probit regresyon kullanmıştır.

Berkson (1944) ise Bliss (1934)'den on yıl sonra probit regresyondan yola çıkarak ilk olarak ikili (binary) lojit model kavramını oraya atmıştır.

Hilbe (2009) “Logistic Regression” isimli kitabında ikili (binary) lojit ve ikili (binary) probit regresyon modellerini ayrıntılı bir şekilde incelemiş olup, bu konuda literatüre önemli bir katkıda bulunmuştur.

Midi ve ark. (2010) 2212 kadın denek üzerine yapılan anket çalışmasında çocuk sahibi olmak isteme ve istememe durumları ile yaş, eğitim durumu, yaşadığı yer gibi çeşitli indikatörler arasındaki istatistiksel ilişkileri ikili (binary) lojistik regresyon modeli aracılığı ile incelemişlerdir.

Abdulqader (2017) hepatit hastalık verilerini yanıt değişkenini “enfekte olmuş hastalar” ve “enfekte olmamış hastalar” almak suretiyle ikili (binary) lojistik regresyon modelini kullanmıştır.

Iyit (2018) enerji güvenliği verilerini yanıt değişkeni çok terimli dağılıma sahip GLM yaklaşımı ile modellemiştir.

Oznur ve Iyit (2018) Amerika Birleşik Devletleri'nde (ABD) şeker hastalığı (diyabet) kaynaklı ölüm oranları ile obezite, çeşitli kanser türleri, şeker tüketimi gibi değişkenler arasındaki istatistiksel ilişkileri Tweedie dağılımına sahip GLM yaklaşımı ile modellemiştir.

Ayan ve Değirmenci (2021) 2017 yılına ait Türkiye İstatistik Kurumu'ndan edindikleri veriler üzerinde tabakalı örnekleme metodu uygulayarak 761 gözlem

üzerinde hane halklarının yoksul olup olmama durumunu çeşitli indikatörler aracılığı ile ikili probit regresyon yaklaşımı ile modellemişlerdir.

Kantarmaci ve ark. (2021) Türkiye SGK'dan edindikleri veriler üzerinde bireylerin aşırı eğitilmiş olup olmama durumları ile yaş, medeni durum, tarım, imalat gibi birçok indikatör arasındaki ilişkiyi ikili probit regresyon yardımıyla modellemişlerdir.

İyit (2021) gamma dağılımına sahip GLM yaklaşımı kullanarak yüksek risk grubu ülkelerin ekonomik büyümesini incelemiştir.

Dobson ve Barnett (2018), Hilbe (2009), Fox (2015), Agresti (2015), Madsen ve Thyregod (2010), McCullagh ve Nelder (1989), Myers ve ark. (2012), Olsson (2002), Dunn ve Smyth (2018), Collett (2003), Faraway (2016) ve Gill ve Torres (2019) GLM alanında literatüre önemli katkılar sunmuşlardır.

2.2. COVID-19 Pandemisi, bulaşıcı olmayan hastalıklar (BOH) ve engelliliğe uyarlanmış yaşam yılları (DALY) ile ilgili literatür taraması

Arsang-Jang ve ark. (2021) ülkelerin gelir durumuna göre COVID-19 ölüm oranı ile BOH'lardan kaynaklanan toplam DALY verileri arasındaki istatistiksel ilişkileri çok düzeyli GLM yaklaşımı kullanarak incelemiştir.

Azarpazhooh ve ark. (2020) korelasyon testlerini kullanarak BOH'lardan kaynaklanan DALY verileri ile COVID-19 pandemisi ile ilgili vaka ve ölüm verileri arasındaki ilişkileri incelemiştir. İran'daki bazı klinik verileri kullanarak COVID-19 pandemisinin neden olduğu ölümler arasında çeşitli BOH'ların oranını araştırmışlardır.

Azadnajafabad ve ark. (2021) İran'daki bazı klinik verileri kullanarak COVID-19 pandemisinin neden olduğu ölümler arasında çeşitli BOH'ların oranını istatistiksel olarak incelemiştir.

Sousa ve ark. (2021) Brezilya'da COVID-19 pandemisi ile hastanede yatan çocuklarda çeşitli BOH'lar ve diğer göstergeler açısından ölüm oranları için risk faktörlerini incelemiştir.

Youn ve ark. (2022) COVID-19 pandemisinin, bazı önemli BOH'ları olan kişiler üzerindeki yayılma etkisini incelemiştir.

Nicoletti-Rojas ve ark. (2022) Şili'deki COVID-19 pandemisi sırasında BOH'ları olan katılımcılardan elde edilen anket verilerini beta regresyon ile modellemişlerdir.

Gaur ve ark. (2021) Hindistan'da çok seviyeli regresyon kullanarak COVID-19 pandemisi ve çeşitli BOH'lardan kaynaklanan vakalar ve ölümler arasındaki ilişkileri incelemişlerdir.

Pécout ve ark. (2021) ABD ve Avrupa'da COVID-19 pandemisinde, COVID-19 hastalarını, BOH'ları olan belirli sayıda katılımcılar arasından seçerek, çeşitli kriterlere göre tanımlayıcı istatistiklerle karşılaştırmışlardır.

İyit ve ark. (2023b) CO₂ emisyon verileri ile COVID-19 pandemi verileri arasındaki istatistiksel ilişkileri GLM ve GLMM yaklaşımları ile modellemişlerdir.

İyit ve ark. (2023a) küresel olarak 179 dünya ülkesinin COVID-19 pandemisinden kaynaklanan ölüm verileri ve kanser ile ilişkili olabilecek bütün tümör ve beraberindeki BOH'lardan kaynaklanan DALY verileri arasındaki ilişkileri genelleştirilmiş lineer model (GLM) yaklaşımı ile incelemişlerdir.

İyit ve Sevim (2023) 174 DSÖ üye ülkelerinin COVID-19 pandemisinden kaynaklanan ölüm verileri ile çeşitli hava kirliliği verileri arasındaki ilişkiyi geometrik, Poisson ve negatif binom (NB) regresyon ve geometrik, Poisson ve negatif binom (NB) karma regresyon modelleri aracılığı ile incelemişlerdir.

Mistry ve ark. (2021) Bangladeş'de 1032 katılımcı üzerinde yapılan anket çalışmasında COVID-19 pandemisi sırasında ilaç ve gerekli tıbbi ihtiyaçlara ulaşımın zor olup olmadığı ile bazı BOH'lar arasındaki ilişkileri ikili lojistik regresyon modeli yaklaşımı ile incelemişlerdir.

Gutierrez ve Bertozzi (2020), Cobre ve ark. (2022), Formenti ve ark. (2022), Azzouzi ve ark. (2022), Yadav ve ark. (2020), COVID-19 pandemisinin kanser ile ilişkili olabilecek bütün tümör ve beraberindeki bulaşıcı olmayan hastalıklardan (BOH) kaynaklanan DALY verileri üzerindeki etkilerini araştırmaya yönelik çalışmalar yapmışlardır.

3. MATERYAL VE YÖNTEM

3.1. Çok değişkenli lineer regresyon modeli

$i = 1, 2, \dots, n$ ve $j = 1, 2, \dots, k$ olmak üzere, y_i : gözlem değerleri (yanıt değişkeni), x_{ij} : k – adet açıklayıcı (bağımsız) değişkenin aldığı değerler, β_0 : kesen(sabit) terim, β_j : kısmi regresyon katsayıları, ε_i : rasgele hata terimleri olacak şekilde i – inci gözlem değeri için çok değişkenli lineer regresyon modeli aşağıdaki gibi yazılabilir (Freund ve ark., 2006; Montgomery ve ark., 2021);

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (3.1)$$

\tilde{y} : yanıt değişkeni vektörü; \mathbf{X} : açıklayıcı (bağımsız) değişkenlere ilişkin tasarım matrisi; $\tilde{\beta}$: bilinmeyen parametre vektörü ve $\tilde{\varepsilon}$: rasgele hata terimleri vektörü olmak üzere;

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}; \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n \times (k+1)}; \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1} \text{ ve } \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \quad (3.2)$$

şeklinde dir. (3.1) eşitliğinin matris halinde gösterimi ise;

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \tilde{\varepsilon} \quad (3.3)$$

(3.1) ya da (3.3) eşitlikleri ile verilen çok değişkenli lineer regresyon modeli için temel varsayımlar aşağıdaki gibi sıralanabilir (Cameron ve Trivedi, 2013; Gamgam ve Altunkaynak, 2015; Fox, 2019);

- ✓ $\text{rank}(\mathbf{X}) = k + 1$ (\mathbf{X} tasarım matrisi tam ranklı olmalıdır)
- ✓ $E(\varepsilon_i) = 0$ (Hataların beklenen değeri sıfırdır)
- ✓ $\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2$ (Hatalar sabit ve eşit varyanslıdır)

- ✓ $i \neq j$ için $Cov(\varepsilon_i, \varepsilon_j) = 0$ (Hatalar birbiri ile ilişkisizdir)
- ✓ $Cov(x_i, \varepsilon_i) = 0$ (Açıklayıcı değişkenlerin rasgele değişken olması durumunda, hatalardan bağımsız olmalıdırlar)
- ✓ $\varepsilon \sim N(0, \sigma^2 I)$ (Güven aralıkları ve hipotez testleri için yani çok değişkenli lineer regresyon modelinde istatistiksel sonuç çıkarımı için hataların çok değişkenli normal dağılıma sahip olması gerekmektedir)

3.2. Genelleştirilmiş lineer modeller

(3.1) ya da (3.3) eşitlikleri ile verilen çok değişkenli lineer regresyon modelinde hataların (güven aralıkları ve hipotez testleri için) normal dağılıma sahip olduğu varsayımı altında yanıt değişkeninin de normal dağılımlı olduğu varsayılır (Gamgam ve Altunkaynak, 2015).

Nelder ve Wedderburn (1972) tarafından ortaya atılan genelleştirilmiş lineer model (GLM) yaklaşımı ile basit ya da çoklu lineer regresyon modelindeki yanıt değişkeninin normal dağılıma sahip olma varsayımı genişletilmiştir.

GLM, yapısal olarak 3 kısımdan oluşmaktadır: yanıt değişkeninin dağılımı, sistematik bileşen ve link (bağlantı fonksiyonu) (Agresti, 2015; Dobson ve Barnett, 2018; Dunn ve Smyth, 2018).

3.2.1. Yanıt değişkenin dağılımı

GLM'de yanıt değişkeninin normal dağılımı da ihtiva eden üstel dağılım ailesinden geldiği varsayılmıştır. Üstel dağılım ailesi olasılık (yoğunluk) fonksiyonu;

$$f(y_i; \phi, \theta_i) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi) \right\} \quad (3.4)$$

olarak tanımlanır. Burada y_i : yanıt değişkeninin gözlenen değerleri, θ_i : kanonik parametre, $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ dağılıma göre bilinen fonksiyonlardır (Myers ve ark., 2012; Gill ve Torres, 2019).

Üstel dağılım ailesine ait bazı dağılımlar; Poisson, binom, negatif binom, geometrik, gamma, normal ve ters normal dağılımlarıdır (Agresti, 2015; İyit ve ark., 2016).

3.2.2. Sistemik bileşen

GLM'de sistemik bileşen (3.1) eşitliği ile verilen çok değişkenli lineer modelde $\beta_0, \beta_1, \dots, \beta_p$ parametrelerinin $x_{i1}, x_{i2}, \dots, x_{ip}$ açıklayıcı değişkenleri ile oluşturduğu lineer kombinasyondur, yani (Fox, 2015; Dobson ve Barnett, 2018);

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (3.5)$$

şeklindedir.

3.2.3. Link (bağlantı) fonksiyonu

GLM'de link fonksiyonu, yanıt değişkenin kendisinden ziyade ortalamasını sistemik bileşen kısmına (lineer kestiricilerin olduğu kısım) bağlamaktadır. $E[y_i] = \mu_i$ ve $g(\cdot)$ link fonksiyonu olmak üzere, GLM'de link fonksiyonu;

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (3.6)$$

olarak gösterilir. (3.4) eşitliği ile verilen üstel dağılım ailesi olasılık (yoğunluk) fonksiyonunda elde edilen $\theta = g(\mu)$ parametresi kanonik (doğal) link fonksiyonu olarak adlandırılmaktadır. GLM'de (3.6) eşitliği aynı zamanda kestirilmek istenen model denklemdir (Agresti, 2015; Faraway, 2016).

GLM'de üstel dağılım ailesine ait bazı dağılımlar ile kullanılan kanonik link fonksiyonları ile (3.6) eşitliğinde verilen $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ fonksiyonları Çizelge 3.1.'de verilmiştir.

Çizelge 3.1. GLM’de sıkça kullanılan bazı link fonksiyonları ve $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ fonksiyonları (Olsson, 2002; Myers ve ark., 2012).

| Dağılım | θ | Kanonik link fonksiyonu $\eta = g(\mu)$ | Ters link fonksiyonu $\mu = g^{-1}(\eta)$ | $a(\cdot)$ | $b(\cdot)$ | $c(\cdot)$ |
|--------------|-----------|--|--|------------|----------------------|---|
| Normal | Birim | μ | η | σ^2 | $\frac{\theta^2}{2}$ | $-\frac{1}{2}\left(\frac{y^2}{\phi} + \ln(2\pi\phi)\right)$ |
| Binom | Lojit | $\ln\left(\frac{\pi}{1-\pi}\right)$ | $\frac{1}{1+\exp(-\eta)}$ | 1 | $n \ln(1+e^\theta)$ | $\ln\left(\frac{n}{y}\right)$ |
| Poisson | Log | $\ln(\mu)$ | $\exp(\eta)$ | 1 | e^θ | $-\ln(y!)$ |
| Gamma, Üstel | Ters | μ^{-1} | η^{-1} | $1/\alpha$ | $-\ln(-\theta)$ | $-\frac{1}{2}\ln\left(\frac{y}{\phi}\right) - \ln\left[y\Gamma\left(\frac{1}{\phi}\right)\right]$ |
| Ters Normal | Ters Kare | μ^{-2} | $\eta^{-1/2}$ | $1/\gamma$ | $-\sqrt{-2\theta}$ | $-\frac{1}{2}\left\{\frac{1}{\phi y} + \ln(2\phi\pi y^3)\right\}$ |

Bu çalışmada yanıt değişkeni binom dağılımlı GLM için kullanılan link fonksiyonları aşağıda verilmiştir (Bliss, 1934; Collett, 2003; Faraway, 2016):

✓ **Lojit link fonksiyonu:**

$$\eta_i = \text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) \quad (3.7)$$

✓ **Probit link fonksiyonu:**

$$\eta_i = \Phi^{-1}(p_i) \quad (3.8)$$

✓ **Tamamlayıcı (complementary) log-log link fonksiyonu:**

$$\eta_i = \log(-\log(1-p_i)) \quad (3.9)$$

✓ **Cauchit link fonksiyonu:**

$$\eta_i = \tan^{-1}(\pi(p_i - 1/2)) \quad (3.10)$$

3.3. Binom dağılımlı yanıt değişkeni için genelleştirilmiş lineer modeller

3.3.1. Lojistik regresyon modeli

p : bir deneyin başarılı olma olasılığı, $1-p$: aynı deneyin başarısız olma olasılığı olmak üzere $y_i \sim \text{Bernoulli}(p_i)$ Bernoulli rasgele değişkeni için olasılık fonksiyonu;

$$f(y_i) = p^{y_i} (1-p)^{1-y_i} ; i=0,1 \quad (3.11)$$

ve olasılık tablosu Çizelge 3.2'de verilmiştir;

Çizelge 3.2. $y_i \sim \text{Bernoulli}(p_i)$ Bernoulli rasgele değişkeni için olasılık tablosu (Myers ve ark., 2012).

| y_i | Olasılık |
|-------|------------------------|
| 1 | $P(y_i = 1) = p_i$ |
| 0 | $P(y_i = 0) = 1 - p_i$ |

y_i yanıt değişkeninin beklenen değeri $E[y_i] = p_i$ olduğundan lojistik regresyon modeli (3.7) eşitliği ile verilen lojit link fonksiyonu aracılığıyla, $i=1,2,\dots,n$ için parametre vektörü $\tilde{\beta} = (\beta_0 \beta_1 \dots \beta_p)'$ ve açıklayıcı değişkenler $\tilde{x}_i = (x_{i1} x_{i2} \dots x_{ip})'$ olmak üzere;

$$\eta_i = g(\mu_i) = \frac{p_i}{1-p_i} = x_i' \tilde{\beta} \quad (3.12)$$

ya da;

$$p_i = E[y_i] = g^{-1}(x_i' \tilde{\beta}) = \frac{\exp(x_i' \tilde{\beta})}{1 + \exp(x_i' \tilde{\beta})} \quad (3.13)$$

olarak yazılır (Collett, 2003; Hilbe, 2009; Madsen ve Thyregod, 2010; Myers ve ark., 2012).

3.3.2. Probit regresyon modeli

Yanıt değişkeni binom dağılımına sahip modellemelerde kullanılan diğer bir regresyon çeşidi de probit regresyon modelidir. $\Phi(\cdot)$ birikimli (kümülatif) standart normal dağılım fonksiyonu;

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{1}{2}u^2\right) du \quad (3.14)$$

şeklinde olmak üzere probit link fonksiyonu $\Phi^{-1}(\cdot)$ ters birikimli standart normal dağılım fonksiyonu olup probit regresyon modeli;

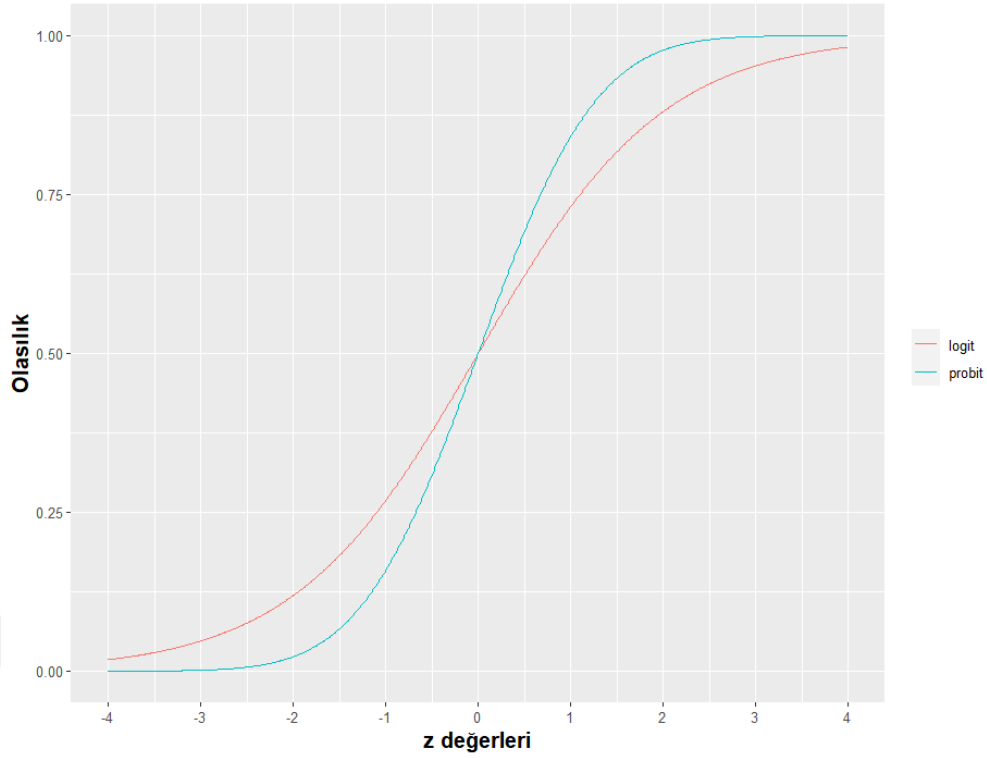
$$\eta_i = g(\mu_i) = \Phi^{-1}(p_i) = x_i' \beta \quad (3.15)$$

ya da;

$$p = \Phi(x_i' \beta) \quad (3.16)$$

olarak yazılır (Bliss, 1934; Collett, 2003; Myers ve ark., 2012).

Yanıt değişkeni binom dağılımına sahip regresyon modellerinde kullanılan lojit ve probit ters link fonksiyonları Şekil 3.1'de grafiksel gösterim olarak verilmiştir.



Şekil 3.1. Yanıt değişkeni binom dağılımına sahip regresyon modellerinde kullanılan lojit ve probit ters link fonksiyonlarının grafik gösterimi

Şekil 3.1’de, x –ekseni $(-4,4)$ aralığında (3.14) eşitliğindeki $\Phi(z)$ ters lojit link fonksiyonundaki z –değerlerine ve (3.16) eşitliğindeki $x_i'\beta$ değerlerine karşılık gelmekte olup y –ekseni ise bu fonksiyonların $(0,1)$ aralığındaki değerlerine karşılık gelmektedir. Şekil 3.1’den görüleceği üzere lojit link fonksiyonu $(0,0.5)$ aralığında probit link fonksiyonuna göre daha fazla değer üretirken, $(0.5,1)$ aralığında ise lojit link fonksiyonu probit link fonksiyonuna göre daha fazla değer üretmektedir.

3.4. Genelleştirilmiş lineer modellerde parametre tahmini

İstatistiksel parametre tahmin yöntemleri olarak literatürde en sık kullanılanlar arasında en küçük kareler yöntemi, en çok olabilirlik yöntemi, momentler yöntemi ve Bayes yöntemi gibi yöntemler gelmektedir (Akdi, 2014).

GLM’de parametre tahmini en çok olabilirlik yöntemine dayanmakta olup log-olabilirlik fonksiyonuna genellikle iteratif yöntemlerle yaklaşım sağlanmaktadır. GLM’de parametre tahminleri Newton-Raphson ve Fisher-scoring ya da IRLS (iteratif

yeniden ağırlıklandırılmış en küçük kareler) yöntemleridir (Olsson, 2002; Hilbe, 2009; 2011).

3.4.1. Genelleştirilmiş lineer modellerde olabilirlik fonksiyonu

En çok olabilirlik yöntemi; tahmini yapılacak parametre ya da parametrelere ait olabilirlik fonksiyonunu maksimum yapan değer ya da değerleri tespit etme esasına dayanmaktadır. Bu nedenle olabilirlik fonksiyonunun logaritması alınarak log-olabilirlik fonksiyonunun ilgili parametre ya da parametrelere göre kısmi türevleri sıfıra eşitlenme suretiyle maksimize olduğu değerler bulunur (Agresti, 2015; Faraway, 2016; Yonar, 2019).

GLM’de üstel dağılım ailesinin (3.4) eşitliği ile verilen olasılık yoğunluk fonksiyonunun olabilirlik ve log-olabilirlik fonksiyonu sırasıyla aşağıdaki gibi yazılır;

$$L(\theta_i, \phi; y_i) = \prod_{i=1}^n f(y_i; \theta_i, \phi) \quad (3.17)$$

$$l(\theta_i, \phi; y_i) = \log \left\{ \prod_{i=1}^n f(y_i; \theta_i, \phi) \right\} = \sum_{i=1}^n \log \{ f(y_i; \theta_i, \phi) \} \quad (3.18)$$

GLM’nin özel bir durumu olan ve bu çalışmanın uygulama kısmında kullanılan lojit regresyon modeli için (3.18) log-olabilirlik fonksiyonunun özel bir hali aşağıdaki gibi elde edilir;

(3.11) eşitliği ile verilen bernoulli dağılımına ait olasılık fonksiyonu, (3.13) eşitliği ile verilen ters lojit link fonksiyonunda yerine yazılırsa;

$$f(y_i) = \frac{\left[\exp(x_i' \beta) \right]^{y_i}}{1 + \exp(x_i' \beta)} \quad (3.19)$$

eşitliği haline gelecektir. (3.19) eşitliği (3.18) log-olabilirlik fonksiyonunda yerine yazıldığında GLM’nin özel bir durumu olan ikili lojit regresyon modelinde log-olabilirlik fonksiyonu;

$$\begin{aligned}
l(\tilde{\beta}) &= \sum_{i=1}^n \log \left\{ \frac{[\exp(x'_i \tilde{\beta})]^{y_i}}{1 + \exp(x'_i \tilde{\beta})} \right\} \\
&= \sum_{i=1}^n \left[y_i (x'_i \tilde{\beta}) - \log \{1 + \exp(x'_i \tilde{\beta})\} \right]
\end{aligned} \tag{3.20}$$

biçiminde yazılır (Collett, 2003; Agresti, 2015).

GLM'nin özel bir durumu olan ve bu çalışmanın uygulama kısmında kullanılan probit regresyon modeli için (3.14) log-olabilirlik fonksiyonunun özel bir hali aşağıdaki gibi elde edilir;

(3.11) eşitliği ile verilen bernoulli dağılımının olasılık fonksiyonu, (3.16) eşitliği ile verilen ters probit link fonksiyonunda yerine yazılırsa;

$$f(y_i) = [\Phi(x'_i \tilde{\beta})]^{y_i} [1 - \Phi(x'_i \tilde{\beta})]^{1-y_i} \tag{3.21}$$

eşitliği haline gelecektir. (3.21) eşitliği (3.18) log-olabilirlik fonksiyonunda yerine yazıldığında GLM'nin özel bir durumu olan ikili lojit regresyon modelinde log-olabilirlik fonksiyonu;

$$\begin{aligned}
l(\tilde{\beta}) &= \log \left\{ \prod_{i=1}^n f(y_i; \theta, \phi) \right\} = \sum_{i=1}^n \log \left\{ [\Phi(x'_i \tilde{\beta})]^{y_i} [1 - \Phi(x'_i \tilde{\beta})]^{1-y_i} \right\} \\
&= \sum_{i=1}^n \left\{ y_i \log [\Phi(x'_i \tilde{\beta})] + (1 - y_i) \log [1 - \Phi(x'_i \tilde{\beta})] \right\} \\
&= \sum_{i=1}^n \left\{ y_i \left(\log [\Phi(x'_i \tilde{\beta})] - \log [1 - \Phi(x'_i \tilde{\beta})] \right) - \log [1 - \Phi(x'_i \tilde{\beta})] \right\}
\end{aligned} \tag{3.22}$$

elde edilir (McCulloch ve Searle, 2001).

3.4.2. Newton-Raphson metodu

Newton-Raphson metodu, GLM'de genel formu (3.18) ile verilen log-olabilirlik fonksiyonuna Taylor seri açılımı yardımı ile iteratif yaklaşım sağlayarak, (3.6) eşitliği ile verilen GLM'de $\beta_0, \beta_1, \dots, \beta_k$ parametre tahminlerini log-olabilirlik fonksiyonun

maksimize olduğu parametre vektörü olarak belirlemektedir (Hardin ve Hilbe, 2007; Müller, 2012).

Öncelikle $l(\underline{\beta})$ log-olabilirlik fonksiyonunun $\beta_0, \beta_1, \dots, \beta_k$ parametrelerine göre birinci dereceden kısmi türev vektörü ile $i = 0, 1, \dots, k$ ve $j = 0, 1, \dots, k$ olmak üzere ikinci derecen kısmi türevlerinin oluşturduğu Hessian matrisi sırasıyla aşağıdaki gibi elde edilir (Myers ve ark., 2012; Agresti, 2015);

$$\underline{u} = \left[\frac{\partial l(\underline{\beta})}{\partial \beta_0}, \frac{\partial l(\underline{\beta})}{\partial \beta_1}, \dots, \frac{\partial l(\underline{\beta})}{\partial \beta_k} \right] \quad (3.23)$$

$$\mathbf{H} = \text{diag} \left(\frac{\partial^2 l(\underline{\beta})}{\partial \beta_i \partial \beta_j} \right) \quad (3.24)$$

$m = 0, 1, 2, \dots$ iterasyon sıra numarası olmak üzere $l(\underline{\beta})$ log-olabilirlik fonksiyonuna $\underline{\beta}^{(m)}$ civarında Taylor seri açılımı uygulanırsa;

$$l(\underline{\beta}) \approx l(\underline{\beta}^{(m)}) + (\underline{u}^{(m)})' (\underline{\beta} - \underline{\beta}^{(m)}) + \frac{1}{2} (\underline{\beta} - \underline{\beta}^{(m)})' \mathbf{H}^{(m)} (\underline{\beta} - \underline{\beta}^{(m)}) \quad (3.25)$$

eşitliği elde edilir (Hilbe, 2009; Gill ve Torres, 2019).

(3.25) iteratif yaklaşımda her iki tarafın $\underline{\beta}$ parametre vektörüne göre türevi alınır;

$$\frac{\partial l(\underline{\beta})}{\partial \underline{\beta}} \approx \underline{u}^{(m)} + \frac{1}{2} \mathbf{H}^{(m)} (\underline{\beta} - \underline{\beta}^{(m)}) + \frac{1}{2} (\underline{\beta} - \underline{\beta}^{(m)})' \mathbf{H}^{(m)} \quad (3.26)$$

eşitliği elde edilir. (3.26) eşitliğinde $\mathbf{H}^{(m)}$ Hessian matrisi (3.24) eşitliğinden de görüleceği üzere köşegen bir matris olduğundan;

$$\mathbf{H}^{(m)}(\tilde{\beta} - \tilde{\beta}^{(m)}) = (\tilde{\beta} - \tilde{\beta}^{(m)})' \mathbf{H}^{(m)} \quad (3.27)$$

eşitliği sağlanacaktır. Dolayısıyla (3.26) iteratif eşitliği $(k+1) \times 1$ boyutlu sıfır vektörüne eşitlenerek;

$$\frac{\partial l(\tilde{\beta})}{\partial \tilde{\beta}} \approx \tilde{u}^{(m)} + \mathbf{H}^{(m)}(\tilde{\beta} - \tilde{\beta}^{(m)}) = \mathbf{0} \quad (3.28)$$

elde edilir. (3.28) eşitliğindeki $\tilde{\beta}$ yalnız bırakılıp bir diğer iterasyondaki alacağı değer

$\tilde{\beta}^{(m+1)}$ olmak üzere;

$$\tilde{\beta}^{(m+1)} = \tilde{\beta}^{(m)} - [\mathbf{H}^{(m)}]^{-1} \tilde{u}^{(m)} \quad (3.29)$$

iteratif eşitliği elde edilir (Härdle ve ark., 1995; Agresti, 2015).

GLM'de log-olabilirlik fonksiyonuna 2. dereceden Taylor serisi açılımı yardımıyla kısmi türevler alınıp sıfıra eşitlenmesi yani en çok olabilirlik (EÇO) tahmini kullanılarak elde edilen (3.29) iteratif eşitliğinde iterasyon, iki iterasyon arasındaki fark daha önceden belirlenen bir değerden küçük olana kadar devam eder (Myers ve ark., 2012; Bakari ve ark., 2016).

3.4.3. Fisher-Scoring metodu

Fisher-scoring metodu, log-olabilirlik fonksiyonunun parametrelere göre türevlerinin sıfıra eşitleme suretiyle oluşturulan denklemlere yaklaşımda Newton-Rahson metoduna alternatif bir yaklaşım metodudur. Newton-Raphson metodundan ayrılan özelliği, Hessian matrisi yerine onun beklenen değerinin negatif hali olan Fisher bilgi matrisinin kullanılmasıdır (Hardin ve Hilbe, 2007; Fox, 2015).

IRLS (iteratively reweighted least squares, iteratif olarak yeniden ağırlıklandırılmış en küçük kareler) metodunda, Newton-Raphson metodunda kullanılan Hessian matrisinin negatif halinin beklenen değeri kullanılmakta olup iteratif metodunda ağırlıklandırma yapılmaktadır (Myers ve ark., 2012; Agresti, 2015).

(3.18) eşitliği ile verilen $l(\theta_i, \phi; y_i) = l(\theta_i)$ olmak üzere üstel dağılım ailesinin log-olabilirlik fonksiyonuna ilişkin Fisher bilgi kriterine özelliklerine dayalı bazı eşitlikler aşağıdaki gibi verilebilir (Rissanen, 1996; Fox, 2015);

$$E \left[\frac{\partial l(\theta_i)}{\partial \theta_i} \right] = 0 \quad (3.30)$$

$$E \left[\frac{\partial^2 l(\theta_i)}{\partial \theta_i^2} \right] + E \left[\left\{ \frac{\partial l(\theta_i)}{\partial \theta_i} \right\}^2 \right] = 0 \quad (3.31)$$

(3.30) özelliği (3.18) log-olabilirlik fonksiyonuna uygulanırsa;

$$E \left[\frac{\partial l(\theta_i)}{\partial \theta_i} \right] = E \left[\frac{y_i - b'(\theta_i)}{a(\phi)} \right] = \frac{E[y_i] - b'(\theta_i)}{a(\phi)} = 0 \quad (3.32)$$

$$\Rightarrow E[y_i] = \mu_i = b'(\theta_i)$$

ve (3.18) ve (3.32) eşitliğinden ;

$$\begin{aligned} & E \left[\frac{\partial}{\partial \theta_i} \left(\frac{y_i - b'(\theta_i)}{a(\phi)} \right) \right] + E \left[\left\{ \frac{y_i - b'(\theta_i)}{a(\phi)} \right\}^2 \right] \\ &= E \left[\frac{-b''(\theta_i)}{a(\phi)} \right] + E \left[\frac{y_i^2 - 2y_i b'(\theta_i) + (b'(\theta_i))^2}{a^2(\phi)} \right] \\ &= \frac{-a(\phi)b''(\theta_i) + E[y_i^2] - 2b'(\theta_i)E[y_i] + (b'(\theta_i))^2}{a^2(\phi)} = 0 \end{aligned} \quad (3.33)$$

$$\Rightarrow E[y_i^2] - (E[y_i])^2 = \text{Var}[y_i] = a(\phi)b''(\theta_i)$$

şeklinde üstel dağılım ailesi için ortalama ve varyans elde edilir (Olsson, 2002; Myers ve ark., 2012).

Üstel dağılım ailesinin eşitlik (3.4) ile verilen formu, (3.18) ile verilen log-olabilirlik fonksiyonunda yerine yazılırsa;

$$l(\theta_i, \phi; y_i) = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi) \right\} \quad (3.34)$$

eşitliği elde edilir. (3.34) eşitliğinin $j=0,1,\dots,k$ olmak üzere β_j parametresine zincir kuralı ile kısmi türevi alınırsa;

$$\frac{\partial l(\theta_i, \phi; y_i)}{\partial \beta_j} = \left(\frac{\partial l(\theta_i, \phi; y_i)}{\partial \theta_i} \right) \left(\frac{\partial \theta_i}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{\partial \eta_i}{\partial \beta_j} \right) \quad (3.35)$$

eşitliği elde edilir. (3.35) eşitliğinin sağ tarafındaki kısmi türevler sırasıyla aşağıdaki gibi elde edilir;

➤ (3.34) eşitliğinin θ_i 'ye göre kısmi türevi;

$$\frac{\partial l(\theta_i, \phi; y_i)}{\partial \theta_i} = \sum_{i=1}^n \frac{y_i - b'(\theta_i)}{a(\phi)} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)} \quad (3.36)$$

➤ (3.32) eşitliğinden $v(\mu_i) = b''(\theta_i)$ (varyans fonksiyonu) olacak şekilde;

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} = \frac{1}{\frac{db'(\theta_i)}{d\theta_i}} = \frac{1}{b''(\theta_i)} = \frac{1}{v(\mu_i)} \quad (3.37)$$

➤ (3.5) eşitliğinden $\eta_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik}$ olduğundan $j=1,2,\dots,p$ olmak üzere;

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \quad (3.38)$$

(3.36), (3.37) ve (3.38) eşitlikleri (3.35) eşitliğinde yerine yazıldığında;

$$\frac{\partial l(\theta_i, \phi; y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{a(\phi) v(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i \quad (3.39)$$

elde edilir. Link fonksiyonu $g(\mu_i) = \theta_i = \eta_i$ olacak şekilde kanonik link fonksiyonu ise;

$$\left(\frac{\partial \mu}{\partial \eta} \right)_i = \left(\frac{\partial \mu}{\partial \theta} \right)_i = b''(\theta_i) \quad (3.40)$$

olarak yazılabileceğinden, (3.39) eşitliği;

$$\frac{\partial l(\theta_i, \phi; y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{a(\phi)} \quad (3.41)$$

olarak ifade edilir (Hardin ve Hilbe, 2007; Agresti, 2015).

Fisher bilgi matrisi olarak da bilinen Hessian matrisinin beklenen değerinin negatif hali aşağıdaki gibi yazılır (Hardin ve Hilbe, 2007; Fox, 2015);

$$\mathbf{I} = -E \left[\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j} \right] = E \left[\left(\frac{\partial l(\beta)}{\partial \beta_i} \right) \left(\frac{\partial l(\beta)}{\partial \beta_j} \right) \right] \quad (3.42)$$

(3.42) eşitliği (3.39) eşitliği ile verilen log-olabilirlik fonksiyonun türevi için uygulanırsa;

$$\begin{aligned} \mathbf{I} &= -E \left[\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j} \right] = E \left[\left(\frac{\partial l(\beta)}{\partial \beta_i} \right) \left(\frac{\partial l(\beta)}{\partial \beta_j} \right) \right] \\ &= E \left(\frac{y_i - \mu_i}{a_i(\phi) v(\mu_i)} \right)^2 \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 x_{ij} x_{ik} \\ &= \frac{(y_i - \mu_i)^2 x_{ij} x_{ik}}{\{a_i(\phi) v(\mu_i)\}^2} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \end{aligned} \quad (3.43)$$

eşitliği elde edilir. $Var(y_i) = (y_i - \mu_i)^2 = a_i(\phi)v(\mu_i)$ olduğundan (3.43) eşitliğinde yerine yazılırsa **I**; Fisher bilgi matrisinin en sade hali ile aşağıdaki gibi ifade edilebilir;

$$\mathbf{I} = \frac{x_{ij}x_{ik}}{Var(y_i)} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \quad (3.44)$$

(3.29) eşitliğinde **H**: Hessian matrisinin yerine beklenen değerinin negatifi olan **I**: Fisher bilgi matrisi, \underline{u} vektörü yerine de (3.39) eşitliğinde üstel dağılım ailesi olasılık yoğunluk fonksiyonu aracılığıyla elde edilen log-olabilirlik fonksiyonu yazılırsa;

$$\tilde{\beta}^{(m+1)} = \tilde{\beta}^{(m)} - \left[\frac{x_{ij}x_{ik}}{Var(y_i)} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \right]^{-1} \left[\frac{(y_i - \mu_i)x_{ij}}{a(\phi)v(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i \right] \quad (3.45)$$

eşitliği elde edilir. Ağırlık olarak adlandırılan matris;

$$\mathbf{W} = \frac{1}{Var(y_i)} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \quad (3.46)$$

biçiminde tanımlandıktan sonra (3.45) eşitliğinin her iki tarafı (3.29) eşitliği ile verilen **I**: Fisher bilgi matrisi ile çarpılıp (3.46) ağırlığı yerine yazılır ve $(m+1)$ -inci iterasyona m -inci iterasyon atanırsa oluşacak eşitliğin sol tarafı;

$$\left[\frac{x_{ij}x_{ik}}{Var(y_i)} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \right] \tilde{\beta}^{(m)} = [\mathbf{X}'\mathbf{W}\mathbf{X}] \tilde{\beta}^{(m)} \quad (3.47)$$

şeklinde olacaktır. (3.46) eşitliğinde **W** matrisi ile $Var(y_i)$ terimleri yer değiştirirse;

$$Var(y_i) = \frac{1}{\mathbf{W}^{-1}} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \quad (3.48)$$

eşitliği elde edilir. (3.48) eşitliği (3.44) eşitliğinde yerine yazılır ve bu eşitlik $\eta_i = x_{ik} \tilde{\beta}^{(m-1)}$ olacak şekilde düzenlenirse;

$$\left[\frac{x_{ij}x_{ik}}{\mathbf{W}^{-1} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2} \left\{ \left(\frac{\partial \mu}{\partial \eta} \right)_i \right\}^2 \right] \tilde{\beta}^{(m-1)} = \mathbf{X}'\mathbf{W}\eta_i \quad (3.49)$$

eşitliği elde edilir. (3.47) ve (3.49) eşitlikleri nihai olarak (3.45) iteratif denkleminde yerine yazılırsa;

$$[\mathbf{X}'\mathbf{W}\mathbf{X}] \tilde{\beta}^{(m)} = \mathbf{X}'\mathbf{W}\eta_i + \left[x_k \mathbf{W} (y_i - \mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i \right] \quad (3.50)$$

eşitliği elde edilir. Çalışma yanıtı (working response) olarak adlandırılan;

$$z_i = \eta_i + (y_i - \mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i \quad (3.51)$$

eşitliği (3.50) eşitliğinde yerine yazılır ve düzenlenirse IRLS iteratif denklemi;

$$\tilde{\beta}^{(m)} = [\mathbf{X}'\mathbf{W}\mathbf{X}]^{-1} \mathbf{X}'\mathbf{W}\tilde{z} \quad (3.52)$$

biçiminde elde edilir (Hardin ve Hilbe, 2007; Stroup, 2012).

3.5. Genelleştirilmiş lineer modellerde bilgi kriterleri

GLM'de kurulan modeller arasında istatistiksel olarak karşılaştırma; bilgi kriterlerine göre yapılmaktadır. Uyum iyiliği açısından istatistiksel olarak en anlamlı model; bilgi kriteri değeri en düşük olan modeldir (McCullagh ve Nelder, 1989; Fox, 2015; Gill ve Torres, 2019; İyit ve ark., 2023b).

p : modelde bulunan parametre sayısı, $l(\theta)$: log-olabilirlik değeri ve n modele alınan toplam gözlem sayısı olacak şekilde GLM’de kullanılan başlıca bilgi kriterlerine ait eşitlikler aşağıdaki gibi verilmiştir;

- ✓ **AIC (Akaike bilgi kriteri)** (Akaike, 1974): $-2l(\hat{\theta}) + 2p$
- ✓ **BIC (Bayesci bilgi kriteri)** (Cavanaugh, 1997): $-2l(\hat{\theta}) + \frac{2pn}{n-p-1}$
- ✓ **AICc (Düzeltilmiş Akaike bilgi kriteri)** (Schwarz, 1978): $-2l(\hat{\theta}) + p \ln(n)$
- ✓ **CAIC (Tutarlı Akaike bilgi kriteri)** (Bozdogan, 1987): $-2l(\hat{\theta}) + (p \ln(n) + 1)$

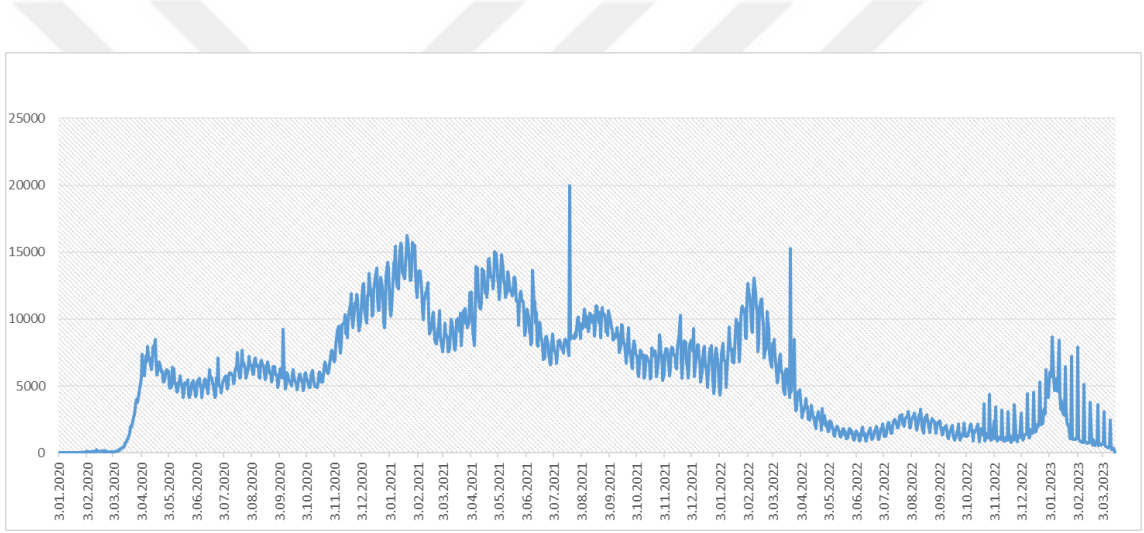


4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Araştırma Sonuçları

Bu çalışmada uygulama olarak 179 dünya ülkesine ait COVID-19 pandemisi verileri ile kanser ve ilişkili olabilecek bütün tümör ve beraberindeki bulaşıcı olmayan hastalıklardan (BOH) kaynaklanan veriler arasındaki ilişkiler genelleştirilmiş lineer model (GLM) yaklaşımı ile incelenmiştir.

03.01.2020 tarihinden 16.04.2023 tarihine kadar COVID-19 pandemisinin sebep olduğu 179 ülkenin günlük toplam ölüm sayılarının çizgi grafiği Şekil 4.1’de verilmiştir (Mathieu ve ark., 2020; İyit ve ark., 2023a).



Şekil 4.1. 179 ülkeye ait COVID-19 pandemisinin sebep olduğu günlük toplam ölüm sayılarının çizgi grafiği

Şekil 4.1’den de görüleceği üzere 03.01.2020 ile 16.04.2023 tarihleri arasında COVID-19 pandemisinin sebep olduğu 179 ülkenin günlük toplam ölüm sayıları 19.991 ile zirve yaptığı gün 21.07.2021 tarihi olarak belirlenmiştir. Bu yüzden bu çalışmaya yanıt değişkeni olarak COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihine kadar 1.000.000 kişiye düşen toplam ölüm sayıları alınmış olup, Çizelge 4.1’de de görüleceği üzere 314.475 medyan değerine göre 0 ve 1 olarak kategorik hale getirilmiştir.

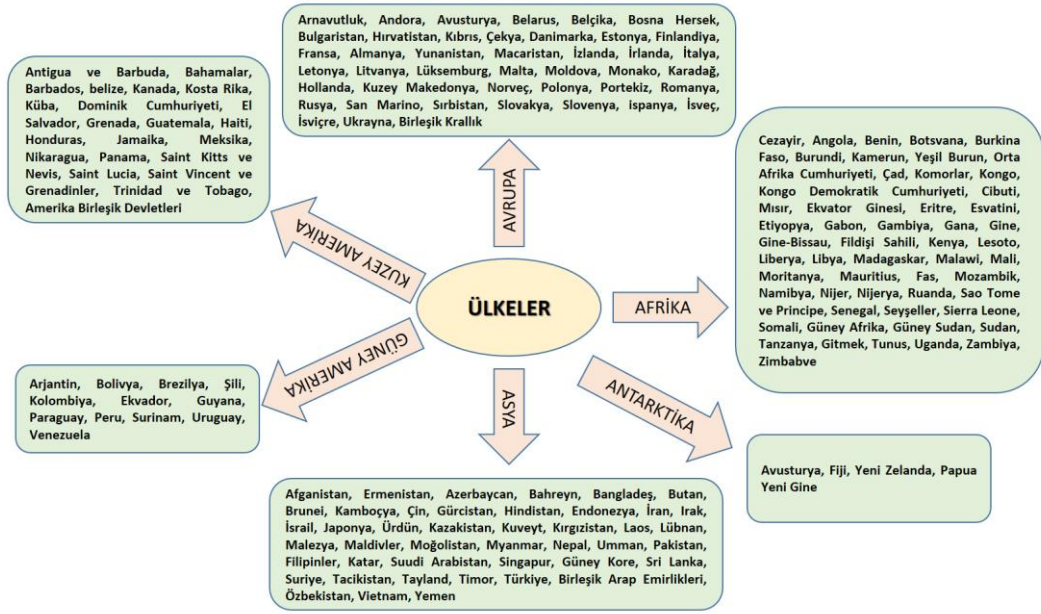
Açıklayıcı değişkenler olarak 179 dünya ülkesinin 2019 yılına ait kanser ile ilişkili olabilecek tümör ve beraberindeki BOH'lardan kaynaklanan 100.000 kişiye düşen DALY verileri kategorik hale getirilerek kullanmıştır. Tez çalışmasında ele alınan

BOH'lar başlıca “siroz ve diğer kronik karaciğer hastalıkları”, “kalp ve damar hastalıkları”, “deri ve deri altı hastalıkları”, “tümör” ile “diğer bulaşıcı olmayan hastalıklar” ’dır. Bir birim DALY, bir yıl sağlıklı yaşam kaybına karşılık gelmektedir. 100.000 kişiye düşen bu BOH'lardan kaynaklanan DALY verileri %25 (1.çeyrek), %50 (2.çeyrek) ve %75 (3.çeyrek) değerlerine göre dört kategorik değer haline getirilerek, bu çalışmaya açıklayıcı değişkenler olarak dâhil edilmiştir.

Çizelge 4.1. Çalışmaya alınan yanıt ve açıklayıcı değişkenlerin çeyrek değerleri ve standart sapmaları

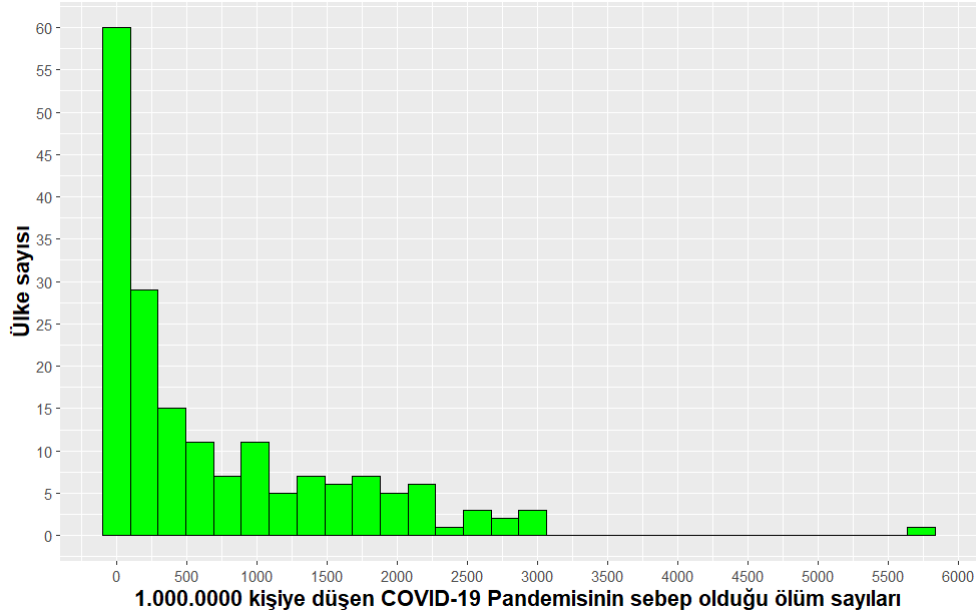
| Çalışmaya Alınan Değişkenler | Min. | 1.çeyrek (%25) | 2.çeyrek (%50-ortanca) | 3.çeyrek (%75) | 4.çeyrek (max) | Std. Sapma |
|--|-------------|-----------------------|-------------------------------|-----------------------|-----------------------|-------------------|
| COVID-19 pandemisinin sebep olduğu toplam ölüm sayıları (100.000 kişiye düşen) | 110.510 | 337.989 | 546.124 | 717.272 | 2616.745 | 360.521 |
| Siroz ve diğer kronik karaciğer hastalıkları | 1367.302 | 2917.532 | 4166.910 | 5614.920 | 18936.290 | 3060.903 |
| Kalp ve damar hastalıkları | 298.863 | 444.101 | 511.050 | 640.589 | 1042.712 | 128.148 |
| Deri ve deri altı hastalıkları | 977.179 | 1872.697 | 2718.874 | 4842.845 | 11701.150 | 1927.764 |
| Tümörler | 1184.196 | 1676.529 | 1911.310 | 2264.387 | 7650.528 | 907.094 |
| Diğer bulaşıcı olmayan hastalıklar | 0.321 | 50.729 | 315.475 | 1122.080 | 5734.078 | 880.738 |

Bu çalışmada uygulama olarak Avrupa kıtasından 43, Kuzey Amerika kıtasından 22, Asya kıtasından 44, Güney Amerika kıtasından 12, Afrika kıtasından 54 ve Antarktika kıtasından 4 olmak üzere toplamda 6 dünya kıtasından 179 ülkenin gözlem değerleri ele alınmıştır. Bu çalışmaya dâhil edilen kıtalar ve ülkeler Şekil 4.2’de gösterilmiştir.



Şekil 4.2. Bu çalışmada ele alınan 6 kıta ve 179 ülkenin şematik olarak gösterimi

179 ülkeye ait 1.000.000 kişiye düşen ölüm sayılarının histogram grafiği Şekil 4.3' de gösterilmiştir.



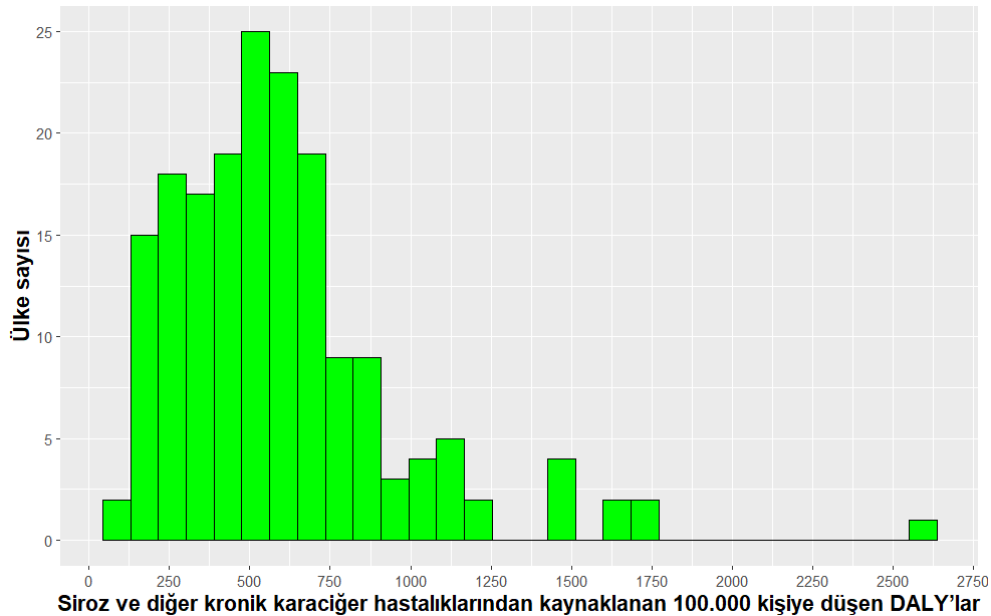
Şekil 4.3. 179 ülkenin 1.000.000 kişiye düşen COVID-19 Pandemisinin neden olduğu ölüm sayılarının histogram grafiği

Dünya genelinde COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayıları en çok olan ilk 5

ülke sırasıyla 5734.078 ile Peru, 3011.747 ile Macaristan, 2990.23 ile Bosna ve Hersek, 2906.425 ile Çekya ve 2681.233 ile Bulgaristan'dır.

Kıta olarak ele alındığında ise COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayıları olarak Avrupa kıtasında en çok ilk 5 ülke 3011.747 ile Macaristan, 2990.23 ile Bosna ve Hersek, 2906.425 ile Çekya, 2681.233 ile Bulgaristan ve 2671.416 ile San Marino; Kuzey Amerika kıtasında en çok ilk 5 ülke 1911.554 ile Meksika, 1791.632 ile Amerika Birleşik Devletleri, 1522.031 ile Panama, 946.951 ile Kosta Rika ve 819.177 ile Belize; Asya kıtasında en çok ilk 5 ülke 1644.685 ile Ermenistan, 1500.38 ile Gürcistan, 1436.679 ile Lübnan, 989.536 ile İran ve 937.349 ile Bahreyn; Güney Amerika kıtasında en çok ilk 5 ülke 5734.078 ile Peru, 2520.771 ile Brezilya, 2422.813 ile Arjantin, 2250.702 ile Kolombiya ve 2114.812 ile Paraguay; Afrika kıtasında en çok ilk 5 ülke 1442.282 ile Tunus, 1129.932 ile Güney Afrika, 1038.167 ile Namibya, 802.726 ile Seyşeller ve 604.986 ile Esvatini; Antarktika kıtasında ise en çok ilk 2 ülke 134.442 ile Fiji ve 35.794 ile Avusturya olmuştur.

179 ülkenin 2019 yılına ait siroz ve diğer kronik karaciğer hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği Şekil 4.4'de verilmiştir.

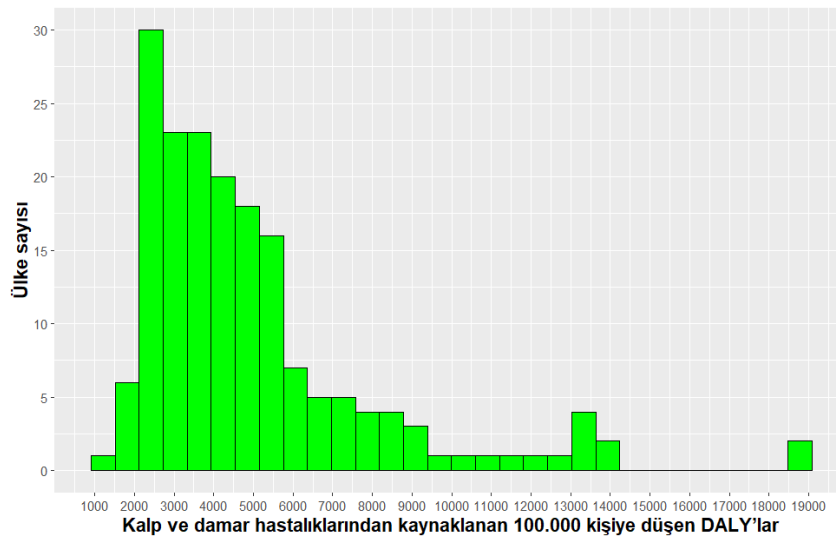


Şekil 4.4. 179 ülkenin 2019 yılına ait siroz ve diğer kronik karaciğer hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği

Dünya genelinde 2019 yılına ait siroz ve diğer kronik karaciğer hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin en çok olduğu ilk 5 ülke sırasıyla 2616.745 ile Moldova, 1743.999 ile Kamboçya, 1718.232 ile Ukrayna, 1675.352 ile Moğolistan ve 1617.875 ile Eritre'dir.

Kıta olarak ele alındığında ise 2019 yılına ait siroz ve diğer kronik karaciğer hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin Avrupa kıtasındaki en çok olduğu ilk 5 ülke 2616.745 ile Moldova, 1718.232 ile Ukrayna, 1454.940 ile Romanya, 1123.535 ile Bulgaristan ve 1106.0554 ile Rusya; Kuzey Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 1119.663 ile Guatemala, 1113.546 ile Meksika, 853.421 ile Honduras; 766.585 ile Saint Kitts ve Nevis ve 762.960 ile Dominik Cumhuriyeti; Asya kıtasında kıtasındaki en çok olduğu ilk 5 ülke 1743.999 ile Kamboçya, 1675.352 ile Moğolistan, 1483.666 ile Mynmar, 1482.692 ile Özbekistan ve 1230.268 ile Gürcistan; Güney Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 1053.979 ile Guyana, 759.148 ile Bolivya, 727.106 ile Şili, 609.682 ile Surinam ve 581.279 ile Ekvador; Afrika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 1617.875 ile Eritre, 1438.986 ile Mısır, 1035.915 ile Seyşeller, 907.531 ile Zambiya ve 873.270 ile Somali; Antarktika kıtasında kıtasındaki en çok olduğu ilk 2 ülke ise 328.154 ile Fiji ve 288.514 ile Papua Yeni Gine olmuştur.

179 dünya ülkesinin 2019 yılına ait kalp ve damar hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği Şekil 4.5'de verilmiştir.

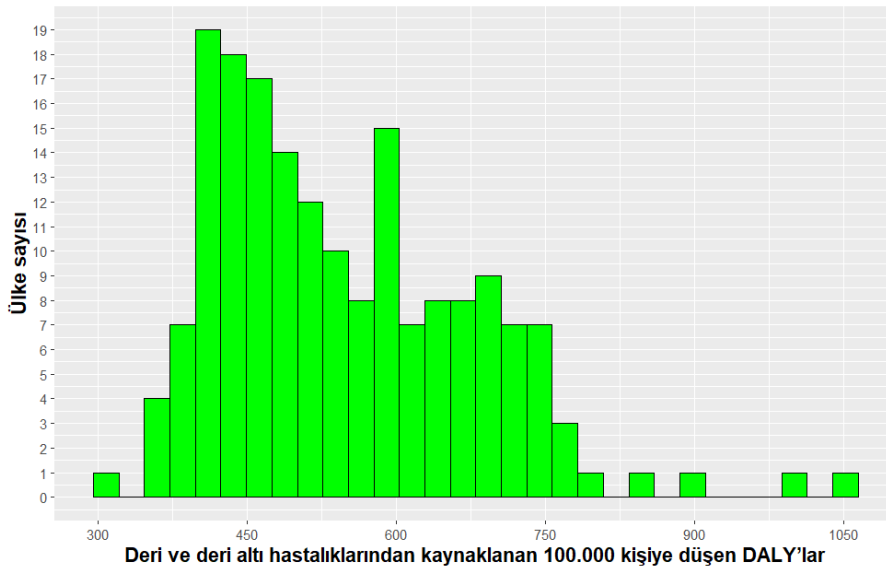


Şekil 4.5. 179 ülkenin 2019 yılına ait kalp ve damar hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği

Dünya genelinde 2019 yılına ait kalp ve damar hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin en çok olduğu ilk 5 ülke sırasıyla 18936.288 ile Bulgaristan, 18785.466 ile Ukrayna, 14133.142 ile Sırbistan, 13962.600 ile Belarus ve 13540.035 ile Moldova'dır.

Kıta olarak ele alındığında ise 2019 yılına ait kalp ve damar hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin Avrupa kıtasındaki en çok olduğu ilk 5 ülke 18936.288 ile Bulgaristan, 18785.466 ile Ukrayna, 14133.142 ile Sırbistan, 13962.600 ile Belarus ve 13540.035 ile Moldova; Kuzey Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 7200.675 ile Saint Kitts ve Nevis, 6736.183 ile Saint Vincent ve Grena, 6433.869 ile Haiti, 6159.147 ile Küba ve 5914.994 ile Barbados; Asya kıtasında kıtasındaki en çok olduğu ilk 5 ülke 13501.038 ile Gürcistan, 9016.083 ile Azerbaycan, 8577.207 ile Ermenistan, 8223.823 ile Özbekistan ve 8180.633 ile Moğolistan; Güney Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 7518.415 ile Guyana, 5396.698 ile Surinam, 4788.884 ile Uruguay, 4406.579 ile Venezuela ve 4184.192 ile Brezilya; Afrika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 7151.113 ile Fas, 6679.323 ile Mısır, 5726.696 ile Tunus, 5722.901 ile Eritre ve 5673.584 ile Mauritius; Antarktika kıtasında kıtasındaki en çok olduğu ilk 2 ülke ise 8093.436 ile Fiji ve 5391.597 ile Papua Yeni Gine olmuştur.

179 dünya ülkesinin 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği Şekil 4.6'da verilmiştir.

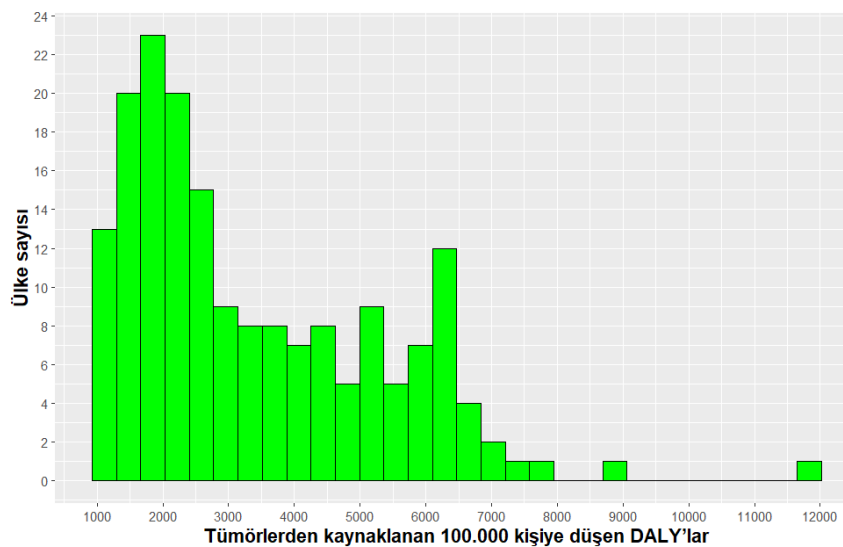


Şekil 4.6. 179 ülkenin 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği

Dünya genelinde 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin en çok olduğu ilk 5 ülke sırasıyla 1042.712 ile Eritre, 1010.161 ile Barbados, 895.861 ile Fransa, 846.333 ile İsrail ve 799.514 ile Saint Kitts ve Nevis'tir.

Kıta olarak ele alındığında ise 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin Avrupa kıtasındaki en çok olduğu ilk 5 ülke 895.861 ile Fransa, 773.819 ile Danimarka, 767.197 ile Portekiz, 761.284 ile Finlandiya ve 754.260 ile Kıbrıs; Kuzey Amerika kıtasındaki en çok olduğu ilk 5 ülke 1010.161 ile Barbados, 799.514 ile Saint Kitts ve Nevis, 687.421 ile Saint Vincent ve Grena, 663.651 ile Haiti ve 658.431 ile Amerika Birleşik Devletleri; Asya kıtasında kıtasındaki en çok olduğu ilk 5 ülke 846.333 ile İsrail, 682.964 ile Timor, 679.201 ile Malezya, 671.844 ile Filipinler ve 648.305 ile Tayland; Güney Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 656.721 ile Arjantin, 654.379 ile Brezilya, 653.449 ile Uruguay, 644.663 ile Paraguay ve 644.487 ile Surinam; Afrika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 1042.712 ile Eritre, 683.683 ile Somali, 614.433 ile Mauritius, 601.837 ile Etiyopya ve 597.609 ile Seyşeller; Antarktika kıtasında kıtasındaki en çok olduğu ilk 2 ülke ise 687.786 ile Papua Yeni Gine ve 681.585 ile Fiji olmuştur.

179 dünya ülkesinin 2019 yılına ait tümörlerden kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği Şekil 4.7'da verilmiştir.

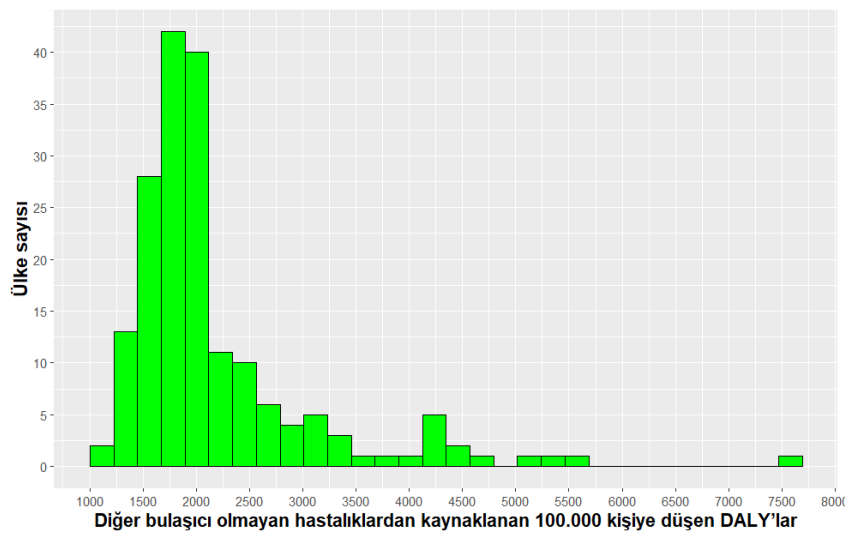


Şekil 4.7. 179 ülkenin 2019 yılına ait tümörlerden kaynaklanan 100.000 kişiye düşen DALY verilerinin ait histogram grafiği

Dünya genelinde 2019 yılına ait tümörlerden kaynaklanan 100.000 kişiye düşen DALY verilerinin en çok olduğu ilk 5 ülke sırasıyla 11701.15 ile Monako, 8917.801 ile Sırbistan, 7927.814 ile Macaristan, 7482.692 ile Bulgaristan ve 7144.312 ile Hırvatistan'dır.

Kıta olarak ele alındığında ise 2019 yılına ait tümörlerden kaynaklanan 100.000 kişiye düşen DALY verilerinin Avrupa kıtasındaki en çok olduğu ilk 5 ülke 11701.15 ile Monako, 8917.801 ile Sırbistan, 7927.814 ile Macaristan, 7482.692 ile Bulgaristan ve 7144.312 ile Hırvatistan; Kuzey Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 6424.532 ile Barbados, 5439.985 ile Küba, 5360.104 ile Saint Kitts ve Nevis, 5177.608 ile Kanada ve 4980.853 ile Amerika Birleşik Devletleri; Asya kıtasında kıtasındaki en çok olduğu ilk 5 ülke 5895.889 ile Japonya, 5674.821 ile Gürcistan, 5287.518 ile Moğolistan, 5156.666 ile Ermenistan ve 4748.684 ile Çin; Güney Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 6157.093 ile Uruguay, 4418.830 ile Arjantin, 3623.336 ile Şili, 3227.166 ile Brezilya ve 3177.933 ile Surinam; Afrika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 4609.785 ile Seyşeller, 3763.186 ile Eritre, 3022.519 Mauritius, 2972.563 ile Lesoto ve 2942.513 ile Yeşil Burun; Antarktika kıtasında kıtasındaki en çok olduğu ilk 2 ülke ise 4508.637 ile Yeni Zelanda ve 4335.632 ile Avustralya olmuştur.

179 dünya ülkesinin 2019 yılına ait diğer bulaşıcı olmayan hastalıklardan kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği Şekil 4.8'da verilmiştir.

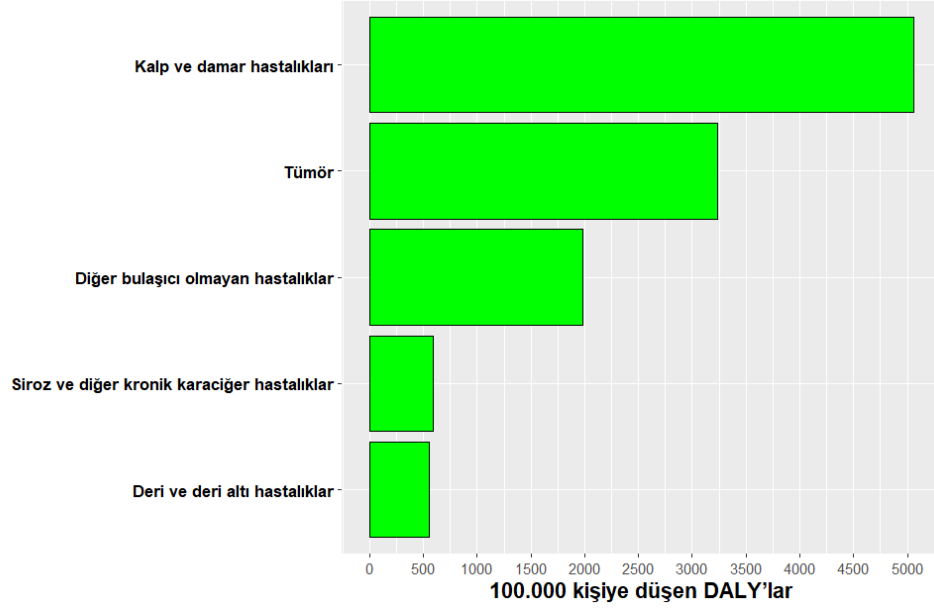


Şekil 4.8. 179 ülkenin 2019 yılına ait diğer bulaşıcı olmayan hastalıklardan kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği

Dünya genelinde 2019 yılına ait diğer bulaşıcı olmayan hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin en çok olduğu ilk 5 ülke sırasıyla 7650.528 ile Burkina Faso, 5556.643 ile Afganistan, 5257.961 ile Mali, 5028.669 ile Sierra Leone ve 4646.642 ile Haiti'dir.

Kıta olarak ele alındığında ise 2019 yılına ait diğer bulaşıcı olmayan hastalıklarından kaynaklanan 100.000 kişiye düşen DALY'ların Avrupa kıtasındaki en çok olduğu ilk 5 ülke 2364.790 ile Avusturya, 2227.287 ile Moldova, 2122.063 ile Belçika, 2112.668 ile Almanya ve 1997.687 ile Danimarka; Kuzey Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 4646.642 ile Haiti, 2354.142 ile Barbados, 2350.028 ile Saint Kitts ve Nevis, 2278.569 ile Saint Vincent ve Grena ve 2015.546 ile Meksika; Asya kıtasında kıtasındaki en çok olduğu ilk 5 ülke 5556.643 ile Afganistan, 3857.620 ile Yemen, 3126.354 ile Tacikistan, 2503.326 ile Ürdün ve 2425.250 ile Pakistan; Güney Amerika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 2649,605 ile Bolivya, 2137.605 ile Brezilya, 2069.957 ile Guyana, 2050.801 ile Uruguay ve 2023.611 ile Surinam; Afrika kıtasında kıtasındaki en çok olduğu ilk 5 ülke 7650.528 ile Burkina Faso, 5257.961 ile Mali, 5028.669 ile Sierra Leone, 4433.120 ile Eritre ve 4391.286 ile Somali; Antarktika kıtasında kıtasındaki en çok olduğu ilk 2 ülke ise 3055.045 ile Papua Yeni Gine ve 1800.388 ile Fiji olmuştur.

BOH'lardan kaynaklanan DALY verileri için dünya geneli incelendiğinde, 2019 yılı için 179 ülke için 100.000 kişiye düşen siroz ve diğer kronik karaciğer hastalıklarından kaynaklanan DALY verileri 594.251, kalp ve damar hastalıklarından kaynaklanan DALY verileri 5058.914, deri ve deri altı hastalıklarından kaynaklanan DALY verileri 552.344, tümörlerden kaynaklanan DALY'lar 3231.694 ve diğer bulaşıcı olmayan hastalıklarından kaynaklanan DALY verileri ise 1978.171 olarak tespit edilmiştir. Şekil 4.9'da 2019 yılı için dünya geneli 100.000 kişiye düşen bu BOH'lardan kaynaklanan DALY verileri için histogram grafiği olarak gösterilmiştir.



Şekil 4.9. 179 ülkenin 2019 yılına ait BOH'lardan kaynaklanan 100.000 kişiye düşen DALY verilerine ait histogram grafiği

Şekil 4.9'dan da görüleceği üzere, 179 dünya ülkesi için yani hemen hemen bütün dünya ülkeleri için incelendiğinde dünyanın en çok sağlıklı olarak yaşam kaybı yaşadığı BOH, 5058.914 DALY yani yaklaşık olarak 100.000 kişide 5059 yıl sağlıklı olarak yitirilmiş yaşam kaybına sebep olan kalp ve damar hastalıkları, ikinci olarak ise yaklaşık olarak 3232 yıl sağlıklı olarak yitirilmiş yaşam kaybına sebep olan tümörlerdir.

4.2. Tartışma

Bu çalışmada, 179 dünya ülkesine ait 21.07.2021 tarihine kadar 1.000.000 kişiye düşen toplam ölüm sayıları ile yine 179 dünya ülkesinin 2019 yılına ait “siroz ve diğer kronik karaciğer hastalıkları”, “kalp ve damar hastalıkları”, “deri ve deri altı hastalıkları”, “tümör” ile “diğer bulaşıcı olmayan hastalıklar” ’dan kaynaklanan 100.000 kişiye düşen DALY verileri medyan değerlerine göre kategorik hale getirilerek aralarındaki istatistiksel ilişki GLM ailesindeki ikili lojit (binary lojit) ve ikili probit (binary probit) regresyon modelleri kullanılarak incelenmiştir. Sonuçlar istatistiksel bilgi kriterlerine (uyum iyiliği istatistikleri) göre karşılaştırılmıştır. Parametre tahmin metodu olarak Fisher-scoring (FS) metodu kullanılmıştır.

Lojit link fonksiyonu altında FS metodu kullanılarak GLM ailesinin özel bir hali olan ikili (binary) lojit regresyon modeli sonuçları Çizelge 4.2’de verilmiştir.

Çizelge 4.2. Lojit link fonksiyonu altında FS metodu kullanılarak GLM ailesinin özel bir hali olan ikili (binary) lojit regresyon modeli sonuçları

| Değişkenler | $\hat{\beta}$ | $s.e(\hat{\beta})$ | Wald Chi-Kare İst. | p – değeri | $\exp(\hat{\beta})$ | 95% Wald güven aralığı ($\exp(\hat{\beta})$ için) | |
|--|---------------|--------------------|--------------------|--------------|---------------------|---|-----------|
| | | | | | | Alt Sınır | Üst Sınır |
| Intercept | 2.6680 | 1.12439 | 5.6304 | 0.01765 * | 14.4113 | 1.5907 | 130.5576 |
| Siroz ve diğer kronik karaciğer hastalıkları [2. seviye] | 0.1002 | 0.89283 | 0.0126 | 0.91066 | 1.1054 | 0.1921 | 6.3606 |
| Siroz ve diğer kronik karaciğer hastalıkları [3. seviye] | 2.1848 | 0.91701 | 5.6763 | 0.01720 * | 8.8887 | 1.4732 | 53.6306 |
| Siroz ve diğer kronik karaciğer hastalıkları [4. seviye] | 0.0405 | 0.82506 | 0.0024 | 0.96086 | 1.0413 | 0.2067 | 5.2467 |
| Kalp ve damar hastalıkları [2. seviye] | 2.6984 | 1.05428 | 6.5508 | 0.01048 * | 14.8558 | 1.8814 | 117.3052 |
| Kalp ve damar hastalıkları [3. seviye] | 1.8255 | 0.95361 | 3.6644 | 0.05559 | 6.2056 | 0.9573 | 40.2269 |
| Kalp ve damar hastalıkları [4. seviye] | 2.8170 | 1.31588 | 4.5829 | 0.03229 * | 16.7268 | 1.2686 | 220.5523 |
| Deri ve deri altı hastalıkları [2. seviye] | 2.7979 | 1.11037 | 6.3493 | 0.01174 * | 16.4101 | 1.8619 | 144.6367 |
| Deri ve deri altı hastalıkları [3. seviye] | 3.8879 | 1.15846 | 11.2631 | 0.00079 *** | 48.8058 | 5.0393 | 472.6861 |
| Deri ve deri altı hastalıkları [4. seviye] | 3.5670 | 1.18812 | 9.0131 | 0.00268 ** | 35.4088 | 3.4496 | 363.4622 |
| Tümörler [2. seviye] | 3.7697 | 0.95914 | 15.4474 | 0.00008 *** | 43.3684 | 6.6181 | 284.1904 |
| Tümörler [3. seviye] | 3.1494 | 1.03589 | 9.2431 | 0.00236 ** | 23.3211 | 3.0618 | 177.6303 |

| | | | | | | | |
|---|-------------|------------|--------|------------|---------------------|--------|---------|
| Tümörler [4. seviye] | 22.091 8 | 1347.79817 | 0.0003 | 0.98692 | 3929743525.45 68 | 0.0000 | Inf |
| Diğer bulaşıcı olmayan hastalıklar [2. seviye] | 2.6862 | 0.96285 | 7.7835 | 0,00527 ** | 14.6764 | 2.2234 | 96.8754 |
| Diğer bulaşıcı olmayan hastalıklar [3. seviye] | 2.4436 | 0.89874 | 7.3926 | 0.00655 ** | 11.5145 | 1.9780 | 67.0300 |
| Diğer bulaşıcı olmayan hastalıklar [4. seviye] | 0.0781 | 0.72107 | 0.0117 | 0.91372 | 1.0813 | 0.2631 | 4.4434 |

FS metodu kullanılarak Çizelge 4.2’de verilen GLM ailesinin özel bir hali olan ikili (binary) lojit regresyon modeli sonuçlarına göre model denklemi;

$$\eta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \left\{ \begin{array}{l} 2.6680 + 2.1848(Karaciğer_hast.[3]) + 2.6984(Kalp_hast.[2]) + \\ 2.8170(Kalp_hast.[4]) + 2.7979(Deri_hast.[2]) + \\ 3.8879(Deri_hast.[3]) + 3.5670(Deri_hast.[4]) + \\ 3.7697(Tümör[2]) + 3.1494(Tümör[3]) + \\ 2.6862(Diğer[2]) + 2.4436(Diğer[3]) \end{array} \right\} \quad (4.1)$$

ya da;

$$\pi_i = \frac{\exp\left\{ \begin{array}{l} 2.6680 + 2.1848(Karaciğer_hast.[3]) + 2.6984(Kalp_hast.[2]) + \\ 2.8170(Kalp_hast.[4]) + 2.7979(Deri_hast.[2]) + \\ 3.8879(Deri_hast.[3]) + 3.5670(Deri_hast.[4]) + \\ 3.7697(Tümör[2]) + 3.1494(Tümör[3]) + \\ 2.6862(Diğer[2]) + 2.4436(Diğer[3]) \end{array} \right\}}{1 + \exp\left\{ \begin{array}{l} 2.6680 + 2.1848(Karaciğer_hast.[3]) + 2.6984(Kalp_hast.[2]) + \\ 2.8170(Kalp_hast.[4]) + 2.7979(Deri_hast.[2]) + \\ 3.8879(Deri_hast.[3]) + 3.5670(Deri_hast.[4]) + \\ 3.7697(Tümör[2]) + 3.1494(Tümör[3]) + \\ 2.6862(Diğer[2]) + 2.4436(Diğer[3]) \end{array} \right\}} \quad (4.2)$$

Probit link fonksiyonu altında FS metodu kullanılarak GLM ailesinin özel bir hali olan ikili (binary) probit regresyon modeli sonuçları Çizelge 4.3’de verilmiştir.

Çizelge 4.3. Probit link fonksiyonu altında FS metodu kullanılarak GLM ailesinin özel bir hali olan ikili (binary) probit regresyon modeli sonuçları

| Değişkenler | $\hat{\beta}$ | $s.e(\hat{\beta})$ | Wald Chi-Kare İst. | $p - \text{değeri}$ | $\exp(\hat{\beta})$ | 95% Wald güven aralığı ($\exp(\hat{\beta})$ için) | |
|--|---------------|--------------------|--------------------|---------------------|---------------------|---|-----------|
| | | | | | | Alt Sınır | Üst Sınır |
| Intercept | 1.4943 | 0.61087 | 5.9837 | 0.01444 * | 4.4562 | 1.3458 | 14.7552 |
| Siroz ve diğer kronik karaciğer hastalıkları [2. seviye] | 0.0054 | 0.50250 | 0.0001 | 0.99137 | 1.0054 | 0.3755 | 2.6921 |
| Siroz ve diğer kronik karaciğer hastalıkları [3. seviye] | 1.1897 | 0.50300 | 5.5942 | 0.01802 * | 3.2861 | 1.2261 | 8.8072 |
| Siroz ve diğer kronik karaciğer hastalıkları [4. seviye] | 0.0685 | 0.47949 | 0.0204 | 0.88634 | 1.0709 | 0.4184 | 2.7410 |
| Kalp ve damar hastalıkları [2. seviye] | 1.4649 | 0.56449 | 6.7344 | 0.00946 ** | 4.3271 | 1.4312 | 13.0826 |
| Kalp ve damar hastalıkları [3. seviye] | 0.9615 | 0.51903 | 3.4319 | 0.06395 | 2.6157 | 0.9457 | 7.2342 |
| Kalp ve damar hastalıkları [4. seviye] | 1.6139 | 0.71629 | 5.0768 | 0.02425 * | 5.0225 | 1.2337 | 20.4475 |
| Deri ve deri altı hastalıkları [2. seviye] | 1.6776 | 0.61043 | 7.5525 | 0.00599 ** | 5.3526 | 1.6179 | 17.7080 |
| Deri ve deri altı hastalıkları [3. seviye] | 2.2692 | 0.63705 | 12.6883 | 0.00037 *** | 9.6716 | 2.7748 | 33.7103 |
| Deri ve deri altı hastalıkları [4. seviye] | 2.0471 | 0.65625 | 9.7308 | 0.00181 ** | 7.7457 | 2.1402 | 28.0333 |
| Tümörler [2. seviye] | 2.1224 | 0.52453 | 16.3720 | 5.2e-05 *** | 8.3508 | 2.9871 | 23.3463 |
| Tümörler [3. seviye] | 1.7429 | 0.56852 | 9.3982 | 0.00217 ** | 5.7138 | 1.8750 | 17.4124 |

| | | | | | | | |
|--|--------|-----------|--------|------------|-----------|--------|---------|
| Tümörler [4. seviye] | 8.2294 | 282.49751 | 0.0008 | 0.97676 | 3749.6396 | 0.0000 | Inf |
| Diğer bulaşıcı olmayan hastalıklar [2. seviye] | 1.5779 | 0.51575 | 9.3604 | 0.00222 ** | 4.8448 | 1.7631 | 13.3135 |
| Diğer bulaşıcı olmayan hastalıklar [3. seviye] | 1.4391 | 0.49531 | 8.4419 | 0.00367 ** | 4.2170 | 1.5973 | 11.1334 |
| Diğer bulaşıcı olmayan hastalıklar [4. seviye] | 0.0482 | 0.42344 | 0.0129 | 0.90942 | 1.0494 | 0.4576 | 2.4064 |

FS metodu kullanılarak Çizelge 4.3’de verilen GLM ailesinin özel bir hali olan ikili (binary) probit regresyon modeli sonuçlarına göre model denklemi aşağıdaki gibi elde edilmiştir.

$$\eta_i = \Phi(\pi_i) = \left\{ \begin{array}{l} 1.4943 + 1.1897(Karaciğer_hast.[3]) + 1.4649(Kalp_hast.[2]) + \\ 1.6139(Kalp_hast.[4]) + 1.6776(Deri_hast.[2]) + \\ 2.2692(Deri_hast.[3]) + 2.0471(Deri_hast.[4]) + \\ 2.1224(Tümör[2]) + 1.7429(Tümör[3]) + \\ 1.5779(Diğer[2]) + 1.4391(Diğer[3]) \end{array} \right\} \quad (4.3)$$

GLM ailesinin özel bir hali olan ikili (binary) lojit ve ikili (binary) probit regresyon modellere ait uyum iyiliği istatistikleri Çizelge 4.4’de verilmiştir.

Çizelge 4.4. Probit link fonksiyonu altında FS metodu kullanılarak GLM ailesinin özel bir hali olan ikili (binary) lojit ve ikili (binary) probit regresyon modellerine ait uyum iyiliği istatistikleri

| Uyum İyiliği İstatistikleri | İkili (Binary) Lojit Model | İkili (Binary) Probit Regression Model |
|-----------------------------|----------------------------|--|
| Log-likelihood | -46.390 | -46.099 * |
| Df. | 16 | 16 |
| Deviance | 92.780 | 92.198 * |
| AIC | 124.780 | 124.198 * |
| AICc | 128.138 | 127.556 * |
| BIC | 175.778 | 175.196 * |
| CAIC | 191.778 | 191.196 * |

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Bu tez çalışmasının sonuç bölümünde, 179 dünya ülkesine ait COVID-19 pandemisi verileri ile kanser ve ilişkili olabilecek bütün tümör ve beraberindeki bulaşıcı olmayan hastalıklardan (BOH) kaynaklanan DALY verileri arasındaki ilişkiyi incelemek üzere oluşturulan GLM ailesinin özel halleri olan ikili (binary) lojit ve ikili (binary) probit regresyon modelleri sonuçları uyum iyiliği istatistiklerine göre karşılaştırılıp en iyi model istatistiksel olarak yorumlanacaktır.

Çizelge 4.4’de görüleceği üzere uyum iyiliği istatistikleri göz önüne alındığında -46.099 ile log-olabilirlik, 92.198 ile sapma (deviance) istatistiği, 124.198 ile AIC, 127.556 ile AICc, 175.196 ile BIC ve 191.196 ile CAIC olmak üzere en küçük değerleriyle GLM ailesinin özel bir hali olan ikili (binary) probit regresyon modeli istatistiksel olarak en iyi model olarak tespit edilmiştir.

Bu çalışmaya özel olarak, hataların istatistiksel anlamda normal bir dağılıma sahip olduğu varsayımına dayanan ikili (binary) probit regresyon modelinin, hataların lojistik bir dağılıma sahip olduğu varsayımına dayanan ikili (binary) lojit regresyon modeline üstünlüğü vurgulanmıştır.

İstatistiksel olarak en iyi model olarak tespit edilen GLM ailesinin özel bir hali olan ikili (binary) probit regresyon modelinde bütün değişkenlerin 1. seviyeleri referans kategori olarak ele alındığında “siroz ve diğer kronik karaciğer hastalıkları” değişkeninin 3. seviyesi, “kalp ve damar hastalıkları” değişkeninin 2. ve 4. seviyeleri, “deri ve deri altı hastalıkları” değişkeninin 2., 3. ve 4. seviyeleri “tümör” değişkeninin 2. ve 3. seviyeleri ve son olarak “diğer bulaşıcı olmayan hastalıklar” değişkeninin 2. ve 3. seviyeleri istatistiksel olarak anlamlı ($p < 0.05$) bulunmuştur.

GLM ailesinin özel bir hali olan ikili (binary) probit regresyon modeline ait Çizelge 4.3’de verilen model sonuçları ve (4.2) eşitliği ile verilen model denklemi göz önünde bulundurulduğunda;

179 dünya ülkesinin 2019 yılına ait siroz ve diğer kronik karaciğer hastalıklarından kaynaklanan 100.000 kişiye düşen DALY’ların 1. seviyesi (110.510 - 337.989 arası) referans alındığında 3. seviyede (546.124 - 717.272) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen

toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(1.1897) = 3.2861$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait kalp ve damar hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (1367.302 - 2917.532 arası) referans alındığında 2. seviyede (2917.532 - 4166.910 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(1.4649) = 4.3271$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait kalp ve damar hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (1367.302 - 2917.532 arası) referans alındığında 4. seviyede (5614.920 - 18936.288 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(1.6139) = 5.0225$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (298.863 - 444.101 arası) referans alındığında 2. seviyede (444.101 - 511.050 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(1.6776) = 5.3526$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (298.863 - 444.101 arası) referans alındığında 3. seviyede (511.050 - 640.589 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(2.2692) = 9.6716$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait deri ve deri altı hastalıklarından kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (298.863 - 444.101 arası) referans alındığında 4. seviyede (640.589 - 1042.712 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(2.0471) = 7.7457$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait tümörlerden kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (977.179 - 1872.697 arası) referans alındığında 2. seviyede (1872.697 - 2718.874 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(2.1224) = 8.3508$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait tümörlerden kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (977.179 - 1872.697 arası) referans alındığında 3. seviyede (2718.874 - 4842.845 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını 1.7429 kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait diğer bulaşıcı olmayan hastalıklardan kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (1184.196 - 1676.529 arası) referans alındığında 2. seviyede (1676.529 - 1911.310 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(1.5779) = 4.8448$ kat arttırmaktadır.

179 dünya ülkesinin 2019 yılına ait diğer bulaşıcı olmayan hastalıklardan kaynaklanan 100.000 kişiye düşen DALY verilerinin 1. seviyesi (1184.196 - 1676.529 arası) referans alındığında 3. seviyede (1911.310 - 2264.397 arası) olması, COVID-19 pandemisinin sebep olduğu 179 ülkenin 21.07.2021 tarihindeki 1.000.000 kişiye düşen toplam ölüm sayılarının kritik değer olarak alınan medyan değerinden (1206.606) büyük olma olasılığını $\exp(1.4391) = 4.2170$ kat arttırmaktadır.

5.2 Öneriler

Her ne kadar daha önce yapılmış çalışmalar mevcut olsa da, COVID-19 pandemisi ile kansere sebep olabilecek tümör ve beraberindeki çeşitli bulaşıcı olmayan hastalıklardan kaynaklanan hastalık yüklerinin modellemelerinde gerek dünya geneli ülke verilerinin kullanılması açısından gerek ise açıklayıcı değişken olarak sadece 5 çeşit bulaşıcı olmayan hastalık verilerinin kullanılması açısından bu çalışma liteartüre özgün katkılar sağlayabilecektir.

Bu çalışma öncülüğünde açıklayıcı değişkenler daha da zenginleştirilip yanıt değişkeni olarak COVID-19 pandemisine ait vaka, test sayısı, aşılama doz sayıları gibi çeşitli veriler kullanılarak GLM ailesine ait diğer dağılımlar ve link fonksiyonları kullanılarak anlamlı modellemeler oluşturulabilir.

Bunlara ek olarak, GLM ailesini de ihtiva eden rasgele etkiler eklenek GLMM'ler (genelleştirilmiş lineer karma modeller) oluşturulup daha yaklaşık sonuçlar elde edilebilir.



KAYNAKLAR

- Abdulqader, Q. M., 2017, Applying the binary logistic regression analysis on the medical data, *Science Journal of University of Zakho*, 5 (4), 330-334.
- Agresti, A., 2015, Foundations of linear and generalized linear models, John Wiley & Sons, p.
- Akaike, H., 1974, A new look at the statistical model identification, *IEEE transactions on automatic control*, 19 (6), 716-723.
- Akdi, Y., 2014, Matematiksel istatistiğe giriş (4. Baskı), Ankara: Gazi Kitabevi, 1-82.
- Akman, M. ve Civek, S., 2022, Dünyada ve Türkiye’de kardiyovasküler hastalıkların sıklığı ve riskin değerlendirilmesi, *The Journal of Turkish Family Physician*, 13 (1), 21-28.
- Anand, S. ve Hanson, K., 1997, Disability-adjusted life years: a critical review, *Journal of health economics*, 16 (6), 685-702.
- Arsang-Jang, S., Belasi, M. T., Najafi, F., Darbandi, M., Raza, M. Z., Akhuanzada, H., Yassi, N., Biller, J., Zand, R. ve Neyra, S. K., 2021, The Association Between Healthcare Resources, Non-communicable Diseases, and Covid-19 Mortality: An Epidemiological Study of 139 Countries.
- Ayan, T. Y. ve Değirmenci, N., 2021, Hanehalkı sorumlusu ve yoksulluk: Probit model, *Journal of Management and Economics Research*, 19 (1), 77-87.
- Azadnajafabad, S., Ghasemi, E., Moghaddam, S. S., Rezaei, N. ve Farzadfar, F., 2021, Non-communicable diseases’ contribution to the COVID-19 mortality: a global warning on the emerging syndemics, *Archives of Iranian Medicine*, 24 (5), 445.
- Azarpazhooh, M. R., Morovatdar, N., Avan, A., Phan, T. G., Divani, A. A., Yassi, N., Stranges, S., Silver, B., Biller, J. ve Belasi, M. T., 2020, COVID-19 pandemic and burden of non-communicable diseases: an ecological study on data of 185 countries, *Journal of Stroke and Cerebrovascular Diseases*, 29 (9), 105089.
- Azzouzi, S., Stratton, C., Muñoz-Velasco, L. P., Wang, K., Fourtassi, M., Hong, B.-Y., Cooper, R., Balikuddembe, J. K., Palomba, A. ve Peterson, M., 2022, The impact of the COVID-19 pandemic on healthy lifestyle behaviors and perceived mental and physical health of people living with non-communicable diseases: An international cross-sectional survey, *International journal of environmental research and public health*, 19 (13), 8023.
- Bakari, H., Adegoke, T. ve Yahya, A., 2016, Application of Newton-Raphson method to non-linear models, *International Journal of Mathematics and Statistics Studies*, 4 (4), 21-31.
- Berkson, J., 1944, Application of the logistic function to bio-assay, *Journal of the American statistical association*, 39 (227), 357-365.
- Bliss, C. I., 1934, The method of probits, *Science*, 79 (2037), 38-39.
- Bourne, D. E., Thompson, M., Brody, L. L., Cotton, M., Draper, B., Laubscher, R., Abdullah, M. F. ve Myers, J. E., 2009, Emergence of a peak in early infant mortality due to HIV/AIDS in South Africa, *Aids*, 23 (1), 101-106.
- Boutayeb, A. ve Boutayeb, S., 2005, The burden of non communicable diseases in developing countries, *International journal for equity in health*, 4, 1-8.
- Bozdoğan, H., 1987, Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, 52 (3), 345-370.
- Cameron, A. C. ve Trivedi, P. K., 2013, Regression analysis of count data, Cambridge university press, p.
- Cavanaugh, J. E., 1997, Unifying the derivations for the Akaike and corrected Akaike information criteria, *Statistics & Probability Letters*, 33 (2), 201-208.

- Chen, A., Jacobsen, K. H., Deshmukh, A. A. ve Cantor, S. B., 2015, The evolution of the disability-adjusted life year (DALY), *Socio-economic planning sciences*, 49, 10-15.
- Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.-C., Wang, C.-B. ve Bernardini, S., 2020, The COVID-19 pandemic, *Critical reviews in clinical laboratory sciences*, 57 (6), 365-388.
- Cobre, A. F., Surek, M., Vilhena, R. O., Böger, B., Fachi, M. M., Momade, D. R., Tonin, F. S., Sarti, F. M. ve Pontarolo, R., 2022, Influence of foods and nutrients on COVID-19 recovery: A multivariate analysis of data from 170 countries using a generalized linear model, *Clinical Nutrition*, 41 (12), 3077-3084.
- Collett, D., 2003, Modeling binary data, *Florida, US: CRC Press*.
- Dobson, A. J. ve Barnett, A. G., 2018, An introduction to generalized linear models, Chapman and Hall/CRC, p.
- Dunn, P. K. ve Smyth, G. K., 2018, Generalized linear models with examples in R, Springer, p.
- Faraway, J. J., 2016, Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models, Chapman and Hall/CRC, p.
- Formenti, B., Gregori, N., Crosato, V., Marchese, V., Tomasoni, L. R. ve Castelli, F., 2022, The impact of COVID-19 on communicable and non-communicable diseases in Africa: a narrative review, *Le Infezioni in Medicina*, 30 (1), 30.
- Fox, J., 2015, Applied regression analysis and generalized linear models, Sage Publications, p.
- Fox, J., 2019, Regression diagnostics: An introduction, Sage publications, p.
- Freund, R. J., Wilson, W. J. ve Sa, P., 2006, Regression analysis, Elsevier, p.
- Gamgam, H. ve Altunkaynak, B., 2015, SPSS uygulamalı regresyon analizi.
- Gaur, K., Khedar, R. S., Mangal, K., Sharma, A. K., Dhamija, R. K. ve Gupta, R., 2021, Macrolevel association of COVID-19 with non-communicable disease risk factors in India, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15 (1), 343-350.
- Geedipally, S. R., Lord, D. ve Dhavala, S. S., 2012, The negative binomial-Lindley generalized linear model: Characteristics and application using crash data, *Accident Analysis & Prevention*, 45, 258-265.
- Gill, J. ve Torres, M., 2019, Generalized Linear Models: A Unified Approach, SAGE Publications, p.
- Gutierrez, J. P. ve Bertozzi, S. M., 2020, Non-communicable diseases and inequalities increase risk of death among COVID-19 patients in Mexico, *PloS one*, 15 (10), e0240394.
- Hardin, J. W. ve Hilbe, J. M., 2007, Generalized linear models and extensions, Stata press, p.
- Härdle, W., Klinke, S., Turlach, B. A., Hilbe, J. ve Turlach, B. A., 1995, Generalized Linear Models, *XploRe: An Interactive Statistical Computing Environment*, 195-222.
- Hilbe, J. M., 2009, Logistic regression models, *Boca Raton*, CRC press, p.
- Hilbe, J. M., 2011, Negative binomial regression, Cambridge University Press, p.
- Horton, R., 2013, Non-communicable diseases: 2015 to 2025, *The Lancet*, 381 (9866), 509-510.
- Islam, S. M. S., Purnat, T. D., Phuong, N. T. A., Mwingira, U., Schacht, K. ve Fröschl, G., 2014, Non-Communicable Diseases (NCDs) in developing countries: a symposium report, *Globalization and health*, 10 (1), 1-8.

- Iyit, N., 2018, Modelling world energy security data from multinomial distribution by generalized linear model under different cumulative link functions, *Open Chemistry*, 16 (1), 377-385.
- İyit, N., Yonar, H. ve Genç, A., 2016, Generalized linear models for European Union countries energy data, *Acta Physica Polonica A*, 130 (1), 397-400.
- İyit, N., 2021, An application of generalized linear model approach on econometric studies, *Research & Reviews in Science and Mathematics-II*.
- İyit, N., Yonar, H. ve Yonar, A., 2021, An application of generalized linear model approach on econometric studies, In: *Research & Reviews in Science and Mathematics-II*, Eds, p. 181-194.
- İyit, N., Sarı, E. ve Sevim, F., 2023a, Modeling COVID-19 Binary Data in the Aspect of Neoplasms as a Potential Indicator of Cancer by Lojtit and Probit Regression Models, *International Journal of Advanced Natural Sciences and Engineering Researches*, 7 (4), 400-407.
- İyit, N. ve Sevim, F., 2023, A novel statistical modeling of air pollution and the COVID-19 pandemic mortality data by Poisson, geometric, and negative binomial regression models with fixed and random effects, *Open Chemistry*, 21 (1), 20230364.
- İyit, N., Sevim, F. ve Kahraman, Ü. M., 2023b, Investigating the impact of CO2 emissions on the COVID-19 pandemic by generalized linear mixed model approach with inverse Gaussian and gamma distributions, *Open Chemistry*, 21 (1), 20220301.
- Kantarmaci, S., Demiroglari, S., Hande, E. ve Üçdoğruk, Ş., 2021, Türkiye İşgücü Piyasasında Bireylerin Aşırı Eğitimliliğinin Ölçülmesi: İki Değişkenli Probit Model Uygulaması, *Sosyoekonomi*, 29 (48), 271-292.
- Lim-Dunham, J. E., Toslak, I. E., Reiter, M. P. ve Martin, B., 2019, Assessment of the American College of Radiology Thyroid Imaging Reporting and Data System for thyroid nodule malignancy risk stratification in a pediatric population, *American Journal of Roentgenology*, 212 (1), 188-194.
- Lone, S. A. ve Ahmad, A., 2020, COVID-19 pandemic—an African perspective, *Emerging Microbes & Infections*.
- Madsen, H. ve Thyregod, P., 2010, Introduction to general and generalized linear models, CRC Press, p.
- Maeda, J. M. ve Nkengasong, J. N., 2021, The puzzle of the COVID-19 pandemic in Africa, *Science*, 371 (6524), 27-28.
- Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D. ve Ortiz-Ospina, E., 2020, Coronavirus pandemic (COVID-19), *Our world in data*.
- McCullagh, P. ve Nelder, J., 1989, Generalized linear models CRC press, In: Boca Raton, Eds: Florida USA, p.
- McCulloch, C. E. ve Searle, S. R., 2001, Generalized, linear, and mixed models (wiley series in probability and statistics).
- Midi, H., Sarkar, S. K. ve Rana, S., 2010, Collinearity diagnostics of binary logistic regression model, *Journal of interdisciplinary mathematics*, 13 (3), 253-267.
- Mistry, S. K., Ali, A. M., Yadav, U. N., Ghimire, S., Hossain, M. B., Das Shuvo, S., Saha, M., Sarwar, S., Nirob, M. M. H. ve Sekaran, V. C., 2021, Older adults with non-communicable chronic conditions and their health care access amid COVID-19 pandemic in Bangladesh: Findings from a cross-sectional study, *PloS one*, 16 (7), e0255534.

- Montgomery, D. C., Peck, E. A. ve Vining, G. G., 2021, Introduction to linear regression analysis, John Wiley & Sons, p.
- Müller, M., 2012, Generalized linear models, *Handbook of Computational Statistics: Concepts and Methods*, 681-709.
- Myers, R. H., Montgomery, D. C., Vining, G. G. ve Robinson, T. J., 2012, Generalized linear models: with applications in engineering and the sciences, John Wiley & Sons, p.
- Nelder, J. A. ve Wedderburn, R. W., 1972, Generalized linear models, *Journal of the Royal Statistical Society: Series A (General)*, 135 (3), 370-384.
- Nicoletti-Rojas, D., Retamal, R., Cerda-Rioseco, R., Rodríguez-Osiac, L., Fuentes-Alburquenque, M. ve Araya-Bannout, M., 2022, Effects of sociodemographic and health factors on the self-management of non-communicable diseases among Chilean adults during the Covid-19 pandemic, *PLOS Global Public Health*, 2 (7), e0000763.
- Olsson, U., 2002, Generalized linear models: an applied approach, Lund, Studentlitteratur, p.
- Oznur, O. ve Iyit, N., 2018, Modelling the US diabetes mortality rates via generalized linear model with the Tweedie distribution, *Int. J. Sci. Res*, 7 (2), 1326-1334.
- Pécourt, C., Pain, E., Chekroun, M., Champeix, C., Kulak, C., Prieto, R., van Vugt, J., Gilchrist, K. ve Lainé-Pellet, A.-F., 2021, Impact of the COVID-19 pandemic on patients affected by non-communicable diseases in Europe and in the USA, *International journal of environmental research and public health*, 18 (13), 6697.
- Rissanen, J. J., 1996, Fisher information and stochastic complexity, *IEEE transactions on information theory*, 42 (1), 40-47.
- Rodríguez, G., 2008, Multilevel generalized linear models, In: *Handbook of multilevel analysis*, Eds: Springer, p. 335-376.
- Sadigh, G., Carlos, R. C., Neal, C. H. ve Dwamena, B. A., 2012, Accuracy of quantitative ultrasound elastography for differentiation of malignant and benign breast abnormalities: a meta-analysis, *Breast cancer research and treatment*, 134, 923-931.
- Schwarz, G., 1978, Estimating the dimension of a model, *The Annals of Statistics*, 461-464.
- Sousa, B. L. A., Brentani, A., Ribeiro, C. C. C., Dolhnikoff, M., Grisi, S. J. F. E., Ferrer, A. P. S. ve Ferraro, A. A., 2021, Non-communicable diseases, sociodemographic vulnerability and the risk of mortality in hospitalised children and adolescents with COVID-19 in Brazil: a cross-sectional observational study, *BMJ open*, 11 (9), e050724.
- Stroup, W. W., 2012, Generalized linear mixed models: modern concepts, methods and applications, CRC press, p.
- Su, X., Yan, X. ve Tsai, C. L., 2012, Linear regression, *Wiley Interdisciplinary Reviews: Computational Statistics*, 4 (3), 275-294.
- Suryasa, I. W., Rodríguez-Gámez, M. ve Koldoris, T., 2021, The COVID-19 pandemic, *International Journal of Health Sciences*, 5 (2).
- Tuttle, K. R., 2020, Impact of the COVID-19 pandemic on clinical research, *Nature Reviews Nephrology*, 16 (10), 562-564.
- Wagner, K.-H. ve Brath, H., 2012, A global view on the development of non communicable diseases, *Preventive medicine*, 54, S38-S41.
- Wedderburn, R. W., 1974, Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method, *Biometrika*, 61 (3), 439-447.

- Weisberg, S., 2005, Applied linear regression, John Wiley & Sons, p.
- WHO, 2019, Disability-adjusted life years (DALYs). WHO; 2019, World Health Organization.
- World_bank, 1993, World development report 1993: investing in health, *Oxford University Press, New York*.
- Yadav, U. N., Rayamajhee, B., Mistry, S. K., Parsekar, S. S. ve Mishra, S. K., 2020, A syndemic perspective on the management of non-communicable diseases amid the COVID-19 pandemic in low-and middle-income countries, *Frontiers in public health*, 8, 508.
- Yonar, H., 2019, Ülkelerin ekonomik büyüme ve risk değerlendirmelerinin geliştirilmiş lineer modeller ile incelenmesi, *Selçuk Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi*.
- Youn, H. M., Quan, J., Mak, I. L., Yu, E. Y. T., Lau, C. S., Ip, M. S. M., Tang, S. C. W., Wong, I. C. K., Lau, K. K. ve Lee, M. S. F., 2022, Long-term spill-over impact of COVID-19 on health and healthcare of people with non-communicable diseases: a study protocol for a population-based cohort and health economic study, *BMJ open*, 12 (8), e063150.
- Zha, L., Lord, D. ve Zou, Y., 2016, The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data, *Journal of Transportation Safety & Security*, 8 (1), 18-35.