

T.C.
OSMANGAZİ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

118 135

KOŞULLU VE SINIRLANDIRILMIŞ LOJİSTİK REGRESYON
YÖNTEMLERİNİN KARŞILAŞTIRILMASI VE BİR UYGULAMA

YÜKSEK LİSANS TEZİ

ERTUĞRUL ÇOLAK

118135

Tez Yöneticisi
Prof. Dr. Kazım ÖZDAMAR

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ
ESKİŞEHİR - 2002

KABUL VE ONAY SAYFASI

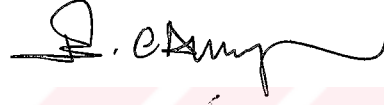
Ertuğrul ÇOLAK'ın Yüksek Lisans Tezi olarak hazırladığı "Koşullu ve Sınırlandırılmış Lojistik Regresyon Yöntemlerinin Karşılaştırılması ve Bir Uygulama" başlıklı bu çalışma, jürimiz Lisansüstü Öğretim Yönetmenliği'nin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

21.03.2002

Prof. Dr. Kazım ÖZDAMAR
ÜYE



Prof. Dr. Ersoy CANKÜYER
ÜYE



Doç. Dr. Alaattin ÜNSAL
ÜYE



Yrd. Doç. Dr. Setenay ÖNER
ÜYE

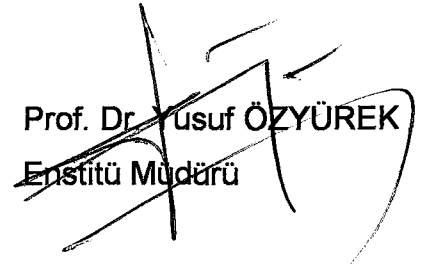


Yrd. Doç. Dr. Fezan ŞAHİN
ÜYE



Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Yönetim Kurulu'nun 21.3.2002 gün ve 53/187 Sayılı kararı ile onaylanmıştır.

Prof. Dr. Yusuf ÖZYÜREK
Enstitü Müdürü



İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER	i
ÖZET	iii
SUMMARY	v
ÇİZELGE DİZİNİ	vii
1. GİRİŞ VE AMAÇ	1
2. GENEL BİLGİLER	5
2.1. Basit ve Çoklu Doğrusal Regresyon Modelleri	5
2.2. Doğrusal Model ve Lojistik Regresyon	9
2.3. İkili Lojistik Regresyon Modeli	10
2.4. Sıralı Lojistik Regresyon Modeli	11
2.5. İsimsel Lojistik Regresyon Modeli	12
2.6. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modelleri	14
2.6.1. Koşullu Lojistik Regresyon Modeli	14
2.6.2. Sınırlandırılmış Lojistik Regresyon Modeli	15
2.7. Parametrelerin Önemliliğinin ve Modelin Uygunluğunun Test Edilmesi	16
2.7.1. Wald Testi	16
2.7.2. Benzerlik Oranı Testi	17
3. GEREÇ VE YÖNTEM	19
3.1. Veri Yapısı	19
3.2. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerinde Kullanılan Bağımlı ve Bağımsız Değişkenler	21
3.3. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerinde Regresyon Katsayılarının Tahminlenmesi	22

3.3.1. Koşullu Lojistik Regresyon Modelinde Regresyon Katsayılarının Tahminlenmesi	23
3.3.2. Sınırlandırılmış Lojistik Regresyon Modelinde Regresyon Katsayılarının Tahminlenmesi	25
4. BULGULAR	32
4.1. Koşullu Lojistik Regresyon Modeli Bulguları	34
4.2. Sınırlandırılmış Lojistik Regresyon Modeli Bulguları	35
5. TARTIŞMA VE SONUÇ	38
KAYNAKLAR DİZİNİ	42
EK	48
ÖZGEÇMİŞ	55



ÖZET

Sağlık alanında karşılaşılan olayların bazıları var-yok, ölü-sağ, hasta-sağlam gibi iki ya da daha çok kategoriler şeklinde sonuçlanmaktadır. Bu sonuçların ortaya çıkmasında bir çok faktör etkili olmaktadır. Kategoriler halinde sonuçlanan olaylar ile bu olayları etkileyen faktörler arasındaki neden sonuç ilişkisini belirlemede lojistik regresyon modelleri kullanılmaktadır.

Sınırlandırılmış ikili veri, gruplar halinde gözlenen ve her grupta en az iki birimin bulunduğu ve her bir grup için en az bir pozitif cevabın elde edildiği veri yapısıdır. Bağımlı değişken olarak sınırlandırılmış ikili veri kullanılan lojistik regresyon modelleri, koşullu ve sınırlandırılmış lojistik regresyon modelleridir. Bu modellerin seçiminde, modele katılan bağımsız değişkenler önemli rol oynamaktadırlar. Eğer modeldeki bağımsız değişkenler birim düzeyinde bağımsız değişkenler ise, her iki modelden biri kullanılabilir. Fakat modele grup düzeyinde bağımsız değişkenler katılmak zorundaysa, sınırlandırılmış lojistik regresyon modeli kullanılmalıdır. Bundan dolayı sınırlandırılmış lojistik regresyon modeli, koşullu lojistik regresyon modeline göre daha etkin bir modeldir.

Bu çalışmada, uygulama olarak Eskişehir Trafik İl Şube Müdürlüğü'nden alınan trafik kazalarına ait veriler kullanılmıştır. Bu veri yapısında önden çarpma sonucu oluşan otomobil kazaları kullanılmıştır. Toplam 20 otomobil ele alınmış ve bu otomobiller grup olarak belirlenmiştir. 20 otomobilde toplam 76 birim yer almaktadır. Önden çarpma sonucu oluşan otomobil kazalarında; emniyet kemeri kullanımı, bireyin yaşı, bireyin otomobildeki oturduğu yer ve otomobilin çarpma hızı ölüm riskini etkileyen faktörler olarak ele alınmış ve bu faktörlerin ölüm riski üzerine etkileri koşullu ve sınırlandırılmış lojistik regresyon modeli kullanılarak araştırılmıştır. Koşullu lojistik regresyon analizi EGRET for WINDOWS Version 2.0.3 paket programı kullanılarak yapılmıştır. Sınırlandırılmış lojistik regresyon analizi ise MATLAB 5.3 programlama dili kullanılarak yazılan bir program ile yapılmıştır.

Koşullu lojistik regresyon analizinde emniyet kemeri kullanımının, bireyin yaşının ve bireyin otomobildeki oturma yerinin ölüm riski üzerine etkileri araştırılmıştır. Otomobilin çarpma hızı otomobil düzeyinde (grup düzeyinde) bağımsız değişken olduğundan koşullu lojistik regresyon analizinde kullanılamamıştır. Sınırlandırılmış lojistik regresyon analizinde ise, emniyet kemeri kullanımının, bireyin yaşının, bireyin otomobildeki oturma yerinin ve otomobilin çarpma hızının ölüm riski üzerine etkileri incelenmiştir.

Her iki modelden de elde edilen analiz sonuçlarına göre; önden çarpma sonucu oluşan otomobil kazalarında emniyet kemeri takmak, otomobilin arka tarafında oturmak ölüm riskini önemli derecede azaltmaktadır. Ancak bireyin yaşının ölüm riski üzerine bir etkisi bulunmamıştır. Otomobilin çarpma hızı sadece sınırlandırılmış lojistik regresyon modelinde analiz edilmiş ve ölüm riski üzerine önemli derecede etkisi olduğu saptanmış ve hız arttıkça ölüm riskinin de arttığı bulunmuştur.

Anahtar Kelimeler: Sınırlandırılmış Kategorik İkili Veri, Koşullu Lojistik Regresyon, Sınırlandırılmış Lojistik Regresyon, Önden Çarpma Sonucu Oluşan Trafik Kazaları, Fisher Information Matrisi, Score Vektörü, Newton-Raphson Algoritması, ODDS Oranı.

SUMMARY

In the area of health some events occurred like that present-absent, death-alive, case-control or in more categories. Occurring of these events affected by many factors. Logistic regression models are used in the determination of cause-effect relation between these categorical events and factors.

Truncated binary data occurs when a group of individuals, who each have a binary response, are observed if one or more of the individuals have a positive response. In conditional and truncated logistic regression models truncated binary data are used. Independent variables are very important in choosing of these models. If the independent variables used in the models are individual level variables, both of the models can be applied. However group level independent variables want to be used in analysis, truncated logistic regression model must be applied. For this reason, truncated logistic regression model is more efficient than conditional logistic regression.

In this research, as an application traffic accident data, which obtained from Eskişehir Traffic Office Department, are applied. These data consist of frontal impact car accidents. Total 20 cars considered and these cars were determined as groups. There are 76 individuals in these 20 cars. In the frontal impact car accidents, usage of safety belt, age of individual, location of individual in the car and collision speed of the car were considered as factors affecting death risk and the effects of these factors on death risk were analyzed by using conditional and truncated logistic regression models. Conditional logistic regression analysis was performed by using package program of EGRET for WINDOWS version 2.0.3. However, truncated logistic regression analysis was performed by using a program written in MATLAB 5.3 programming language.

In the conditional logistic regression analysis, the effects of usage of safety belt, age of individual and location of individual in the car, on death risk were investigated. While collision speed of the car is a group level independent variable, it was not used in the truncated logistic regression analysis. In the truncated logistic regression analysis, the effects of usage of safety belt, age of individual, location of individual in the car and collision speed of the car, on death risk were analyzed.

According to the results obtained from the both models; in the frontal impact car accidents fasten safety belt and locate in the back seat of the car decreased death risk significantly. But the effects of age of individual on death risk didn't find. Because collision speed of the car is a group level independent variables, it analyzed only in the truncated logistic regression model. The effect of collision speed of the car on death risk was found significantly. When the collision speed of the car increases, death risk of individual in the car increases.

Key Words: Truncated Binary Data, Conditional Logistic Regression, Truncated Logistic Regression, Front Impact Car Accidents, Fisher Information Matrix, Newton-Raphson Algorithm, ODDS Ratio.

ÇİZELGE DİZİNİ

Çizelge	Sayfa
4.1. Eskişehir 2000-2001 Yılı Önden Çarpma sonucu Oluşan Otomobil Kazaları Verileri	32
4.2. Uygulamada Kullanılan Sabit ve Bağımsız Değişkenlerin Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerine Katılma Durumu	34
4.3. Koşullu Lojistik Regresyon Modeli Analiz Sonuçları	35
4.4. Sınırlandırılmış Lojistik Regresyon Modeli Analiz Sonuçları	36
4.5. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerinden Elde Edilen Parametre Tahminlerinin Karşılaştırılması	37



1. GİRİŞ ve AMAÇ

Doğada olaylar incelendiğinde olayların detayları anlaşıldıkça, arka planda yatan bir çok nedenin bulunduğu da fark edilmektedir. Sağlık alanındaki oluşumlar; teknolojik gelişmelere ve karşılaşılan problemlere göre daha çok ayrıntılı biçimde incelemeyi gerektiren boyutlara ulaşmış bulunmaktadır. Gün geçtikçe basit, tekdeğişkenli ve ikili ilişkilerle açıklanmaya çalışılan sorunlar, artık daha ayrıntılı biçimde çoklu ya da çokdeğişkenli yöntemlerle açıklanma gereğini gündeme getirmektedir.

Sağlık olaylarına getirilen çözümlerin boyutları günümüzde çok detaylanmış bulunmaktadır. Eskiden bilimsel çalışmalarda yer verilmeyen değişkenler, değişkenlerarası ilişkiler de günümüzdeki araştırmalarda ele alınarak çözümlenmesi hedeflenmektedir.

Sağlık olaylarında, olaylar arasındaki neden sonuç ilişkilerinin ortaya konması, risk faktörlerinin belirlenmesi ve bu risk faktörlerinden hangilerinin majör öneme sahip olduğu hangilerinin minör yapıda, önemsenmeyecek boyutlarda olduğunun belirlenmesi ileri istatistiksel yöntemlerle ortaya konmaya çalışılmaktadır.

Yaygın olarak bilinen (konvansiyonel) Lojistik Regresyon Modelleri ile incelenen sağlık olaylarında bu olayların gözlenme olasılıklarının belirlenmesinde etkili olan değişkenler tümüyle birim bazında ortaya konabilen değişkenler idi. Güncel araştırmaların ortaya koyduğu sonuçlar gösterdi ki, konvansiyonel yöntemlerle incelenemeyen yer, zaman ve birimin içinde bulunduğu gruba ilişkin sabit özellik taşıyan değişkenler de risk faktörleri arasında yer almakta ve bu faktörlerin önemliliğinin değerlendirilmesi gerekmektedir. Konvansiyonel yöntemler değişkenlerin birimlere göre az ya da çok değişkenlik (varyasyon) göstermesini gerektirmektedir. Ancak grup bazında, yersel dağılım bazında ve zaman boyutunda n sayıda birimin ortak maruz

kaldığı değişkenler grup içi değişkenlik göstermemektedir. Grup içi değişkenliğe sahip olmayan veriler ise konvansiyonel yöntemlerle analiz edilememektedir.

Grup, yer, zaman boyutlarında incelenen sağlık olaylarının risk faktörleri yönünden incelenmesi ve herbir faktörün olayın oluşumu (gözlenme/gözlenmeme) üzerine etkilerini ortaya koymak amacıyla Sınırlandırılmış Lojistik Regresyon (Truncated Logistic Regression) ve Koşullu Lojistik Regresyon (Conditional Logistic Regression) yöntemleri yeni geliştirilmiş yöntemler arasında yer almaktadır. Birimlerin içinde bulunduğu yer-zaman ve grup koşullarında, ortak etkilere sahip oldukları konumlarda diğer faktörlerin olayın gözlenme olasılığına etkisinin analizinde yardımcı olan bu yöntemler sağlık alanında sıklıkla karşılaşılan risk faktörlerinin grup düzeyindeki etkilerini değerlendirme bakımından etkin yöntemler arasında yer almaktadır (42, 54, 55).

Günümüzde sağlık alanında yapılan bilimsel araştırmalarda güvenilir sonuçlara varılabilmesi için, kararların istatistik yöntemlere dayandırılması gerekmektedir. Sağlık bilimlerinde, incelenen olayların karmaşık ve bu olayların çözümü için önerilen yolların da fazla olması, olayı açıklamada kullanılacak değişken sayısını da artırmaktadır. Bu amaçla olayların çözümünde, birden fazla değişkeni konu alıp, bunların analizleriyle uğraşan istatistiksel modellerin kullanılması gerekmektedir.

Elde edilen verilerin analiziyle ilgilenildiğinde, kuramsal istatistik bir modelin matematiksel fonksiyonlarla ifade edilmesi gerekmektedir. Çünkü bu fonksiyonlar, gözlenen verilerden ileride gerçekleşmesi muhtemel olaylar hakkında tahmin yapılmasına ve olaylara etki eden faktörlerin belirlenmesine olanak sağlarlar. Olaylara çözüm getirmede aynı amaç için farklı modeller gerekebilir. Fakat, aynı amaç için farklı modellerin kullanılmasının bazı riskleri bulunabilir. Uygulanacak modelin seçiminde izlenecek yol, mevcut riskleri en aza indirgeyen modelin seçilmesidir.

Sağlık alanında karşılaşılan olayların bazıları ölü-sağ, hasta-sağlam, yok-orta-çok, olumsuz-olumlu-çok olumlu ya da daha çok kategoriler biçiminde sonuçlanabilir (43). Bu sonuçların ortaya çıkmasında bir çok faktör rol oynar. Bu sonuçlarla, faktörler arasındaki ilişkinin istatistiksel yöntemlerle belirlenmesinde lojistik regresyon modelleri kullanılır. Bu tip sonuçlar bazı durumlarda, birbirinden bağımsız sınırlandırılmış ya da gruplandırılmış veri setlerinden elde edilmektedir.

Sağlık alanında yapılan çalışmalarda gruplandırılmış veri setlerine sıkça rastlanılmaktadır. Gruplandırılmış verilerde, grup etkilerinin de analize katılması doğru sonuçlara ulaşmada büyük önem taşımaktadır. Ancak, grup etkilerinin analize katılabilmesi için kullanılan tek değişkenli (univariate) istatistiksel yöntemler yetersiz kalmaktadır. Bu nedenle, daha gelişmiş yöntemler olan çokdeğişkenli (multivariate) istatistiksel modeller kullanılmalıdır. Sınırlandırılmış ya da gruplandırılmış veri setlerinden elde edilen kategorik sonuçlarla, bu sonuçların oluşumuna etkide bulunan faktörler arasındaki ilişkinin istatistiksel yöntemlerle incelenmesinde Sınırlandırılmış Lojistik Regresyon ya da Koşullu Lojistik Regresyon modelleri kullanılmaktadır (12, 28). Her iki model de sınırlandırılmış ya da gruplandırılmış veri setlerine uygulanabilirken, sınırlandırılmış lojistik regresyon modelinde grup etkileri incelenebilmektedir. Bu nedenle, bu çalışmanın en önemli noktası sınırlandırılmış lojistik regresyon modelinin, koşullu lojistik regresyon modeline göre daha etkin bir model olduğunu göstermektir.

Tıp alanında elde edilen verilerin çoğu sınırlandırılmış ya da gruplandırılmış kategorik veri yapısına uyduğundan, bu veri yapılarının analizinde uygulanacak yöntemin seçimi doğru sonuçlara ulaşmada büyük önem taşımaktadır.

Gelişmiş ve azgelişmiş ülkelerde trafik kazaları ölüm nedenleri arasında önemli yer tutmaktadır. Özellikle araç sayılarındaki artışa paralel olarak ölümlü kazalarda da önemli artışlar olmaktadır. Kazalarda ölüme neden

olan faktörlerin matematik-istatistik modellerle incelenmesi ve parametre tahminlerinde geçerli güvenilir yaklaşımların uygulanması, araştırmalardan elde edilen sonuçların tutarlılığı bakımından büyük önem taşımaktadır. Trafik kazalarından ölümler; dikkat, uygun yol, uygun hız, emniyet kemeri kullanma, sürücülerin eğitimi ve kontrollü-güvenilir araç kullanımı gibi koşullarla önlenabilir. Bu koşullara yeni koşulların eklenmesi verilerin ileri istatistiksel yöntemler ve bilimsel yaklaşımlarla incelenmesi gerekmektedir (42, 54, 55).

Bu çalışma;

- Sınırlandırılmış ve Koşullu Lojistik Regresyon modellerini açıklamak,
- Bu iki modelin arasındaki temel farkları saptamak,
- Aynı amaçla kullanılan bu modellerin hangi durumda hangi tip veri yapısına uygulanacağını belirlemek,
- Sınırlandırılmış ve Koşullu Lojistik Regresyon modellerini bir uygulama üzerinde göstermek,

amaçlarını gerçekleştirmek üzere yapılmıştır.

2. GENEL BİLGİLER

Lojistik regresyon; bağımlı değişkenin ikili, üçlü ve çoklu kategoriler halinde gözlemlendiği durumlarda, açıklayıcı değişkenlerle neden sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Açıklayıcı değişkenlere göre cevap değişkeninin beklenen değerinin, olasılık olarak elde edildiği bir regresyon yöntemidir (43).

Lojistik regresyon analizinin açıklanmasında kullanılan bazı kavram, gösterimler ve modeller aşağıda kısaca açıklanmıştır.

2.1. Basit ve Çoklu Doğrusal Regresyon Modelleri

Basit ve çoklu doğrusal regresyon analizleri de bağımlı değişken ile açıklayıcı değişken ya da değişkenler arasındaki matematiksel bağıntıyı analiz etmekte kullanılmaktadır.

Bağımlı Değişken : Değeri başka değişkenler tarafından etkilenen değişkene bağımlı değişken denir. İncelenen bir olayda, sonuç değişken bağımlı değişkendir. Bağımlı değişken genellikle Y harfi ile gösterilir.

Bağımsız Değişken : Değeri rasgele koşullara göre belirlenen, bağımsız olarak değişim gösteren ve başka değişkenlerin de değişimi üzerine etkide bulunan değişkelere, bağımsız değişken denir. Bağımsız değişken genel olarak X harfi ile gösterilir. Bağımsız değişkenler ya ortak ya da faktör değişkenlerdir.

Ortak Değişken : Bağımlı değişkenlerle birliktelik içindeki aralıklı ya da oransal ölçekli sürekli değişkenlere ortak değişken denir (25, 40, 43).

Faktör Değişken : Bağımlı değişkenlerle neden-sonuç ilişkisi içindeki kategorik, isimsel ya da sıralı ölçekli değişkenlere faktör değişken denir (43).

Basit doğrusal regresyon modeli;

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

şeklinde ve çoklu doğrusal regresyon modeli ise;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2)$$

şeklinde ifade edilir. Burada;

Y: Bağımlı değişkeni,

X / X₁, X₂, ..., X_k: Bağımsız (Açıklayıcı) değişken / değişkenleri,

β_0 : Bağımsız değişken / değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

ε : Hata terimini,

k: Bağımsız değişken sayısını, göstermektedir.

Basit ve çoklu doğrusal regresyon modelinde, bağımlı değişkenin, verilen bağımsız değişken ya da değişkenlerin değerlerine göre beklenen değeri (ortalama değeri) basit doğrusal regresyon modeli için;

$$E(Y/X) = \beta_0 + \beta_1 X \quad (3)$$

şeklinde ve çoklu doğrusal regresyon modeli için ise;

$$E(Y/X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (4)$$

şeklinde gösterilir. Hata teriminin beklenen değeri sıfırdır. Parametre tahminleri yukarıdaki bağımlı değişkenin beklenen değerlerini veren modellere göre yapılmaktadır (40, 43).

Basit ve çoklu doğrusal regresyon yöntemlerinin uygulanabileceği veri setlerinde;

Bağımlı değişkenin (Y) normal dağılım göstermesi,

Bağımsız değişkenlerin normal dağılım gösteren toplum ya da toplumlardan çekilmiş olması ve hatasız ölçümler olarak belirlenmesi,

Bağımsız değişkenler arasında çoklu bağımlılık (multicollinearity) olmaması,

Hata teriminin, tüm bağımsız değişkenler için sıfır ortalamalı ve σ^2 varyanslı normal dağılım göstermesi, $\varepsilon \cong N(0, \sigma^2)$,

Hata terimleri arasında otokorelasyon olmaması, $E(\varepsilon_i, \varepsilon_j) = 0$,

Hata terimleri ile bağımsız değişkenler arasında bir korelasyon olmaması, $E(\varepsilon_i, X_i) = 0$,

varsayımları gerekmektedir (16, 38, 40).

Yukarıdaki varsayımların yerine getirilemediği veri setlerine basit ya da çoklu doğrusal regresyon analizi uygulanamaz. Lojistik regresyon analizinde normal dağılım ve süreklilik varsayımı ön koşulu yoktur. Bağımlı değişken üzerinde açıklayıcı değişkenlerin etkileri olasılık olarak elde edilir (43).

ODDS Deęeri : İncelenen bir olayın olasılıęının kendi dıřında kalan dięer olayların olasılıęa oranına ODDS deęeri denir ve

$$\text{ODDS}_P = \frac{P}{1-P} \quad (5)$$

řeklinde gsterilir. Burada P incelenen olayın olasılıęını gstermektedir. ODDS deęeri 0 ile $+\infty$ arasında deęerler alır (1, 20, 38).

ODDS Oranı : İncelenen iki farklı olayın ODDS deęerlerinin birbirine oranına ODDS oranı denir. ODDS oranı incelenen iki olayın gzlenme olasılıklarından birinin dięerine oranla ka kat daha fazla veya ka kat daha az olarak ortaya ıkabileceęini gsterir.

Eęer incelenen bir A olayının E kmesi iinde ortaya ıkma olasılıęı;

$$P(A / E) \quad (6)$$

ile gsterilir ise, A olayının E kmesi iindeki ODDS deęeri;

$$\text{ODDS}_{P(A/E)} = \frac{P(A/E)}{1-P(A/E)} = \frac{P(A/E)}{P(\bar{A}/E)} \quad (7)$$

řeklinde ifade edilir. İncelenen A olayının E kmesinin dıřında ortaya ıkma olasılıęının ODDS deęeri ise;

$$\text{ODDS}_{P(A/\bar{E})} = \frac{P(A/\bar{E})}{1-P(A/\bar{E})} = \frac{P(A/\bar{E})}{P(\bar{A}/\bar{E})} \quad (8)$$

řeklinde ifade edilir. Buradan, bu iki ODDS deęeri birbirlerine oranlanır ise;

$$\text{ODDS ORANI} = \frac{P(A/E)/P(\bar{A}/E)}{P(A/\bar{E})/P(\bar{A}/\bar{E})} \quad (9)$$

ifadesi elde edilir. ODDS oranı 0 ile $+\infty$ arasında değerler alır. Eğer bu oran 1'den büyük çıkarsa, incelenen A olayının, E kümesi içinde ortaya çıkma olasılığının E kümesi dışında gözlenme olasılıklarına göre o kadar kat artacağını, eğer 1'den küçük çıkarsa, incelenen A olayının, E kümesi içinde ortaya çıkma olasılığının E kümesi dışında gözlenme olasılığına göre o kadar kat azalacağını gösterir (20, 56).

Logit Fonksiyon : Logit fonksiyon, incelenen bir olasılığın (P), ODDS değerinin doğal logaritmasını verir. İncelenen olasılığın (P) logit fonksiyonundaki gösterimi;

$$\text{Logit}[P] = \ln\left(\frac{P}{1-P}\right) = \ln(\text{ODDS}_p) \quad (10)$$

şeklindedir.

İncelenen bir olasılığın ODDS değeri 0 ile $+\infty$ arasında değer alırken aynı olasılığın logit değeri $-\infty$ ile $+\infty$ arasında değerler alır (1,12).

2.2. Doğrusal Model ve Lojistik Regresyon

Olasılıkların logit fonksiyonunun kullanılması amacını, doğrusal bir model elde edilerek, parametre tahminlerinin yapılmasıdır. Doğrusal modeli;

$$\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (11)$$

şeklinde gösterirsek ve incelenen bir olasılığın (P) logit değerini bu doğrusal modele eşitlersek;

$$\text{Logit}[P] = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (12)$$

eşitliğini elde ederiz. Elde ettiğimiz bu eşitlik, basit ve çoklu doğrusal regresyon modellerindeki bağımlı değişkenin beklenen değerini veren ve parametre tahminlerinde kullanılan eşitliklere benzer bir eşitliktir. Çünkü $E(Y/X_1, X_2, \dots, X_k)$ değeri ile $\text{Logit}[P]$ değeri $-\infty$ ile $+\infty$ arasında değerler almaktadır. Bu eşitlikten incelenen olasılık (P);

$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}} \quad (13)$$

şeklinde elde edilir. Bu eşitliğe lojistik regresyon modeli denir. Burada;

P: incelenen olayın gözlenme olasılığını,

β_0 : Bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

X_1, X_2, \dots, X_k : Bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

$e=2,718281828$ sayısını göstermektedir (20).

2.3. İkili Lojistik Regresyon Modeli (Binary Logistic Regression)

Bağımlı değişkenin ikili kategorik cevap içerdiği lojistik regresyon modelidir. Bağımlı değişken var-yok, hasta-sağlam, başarılı-başarısız gibi değerler alır. Bir ya da daha fazla bağımsız değişken ile ikili cevap değişken arasındaki bağıntıyı ortaya koyar. Bağımsız değişkenler ya faktör değişken ya da ortak değişkenlerdir. Bağımlı değişkende incelediğimiz kategori genel olarak $Y=1$ ile kodlanır. Diğer kategoride $Y=0$ ile ifade edilir. Bu durumda incelediğimiz

kategorinin olasılık deęerini baęımsız deęiřkenlerle analiz eden ikili lojistik regresyon modeli;

$$P(Y_j = 1) = \frac{e^{\beta_0 + \beta_1 X_{j1} + \dots + \beta_k X_{jk}}}{1 + e^{\beta_0 + \beta_1 X_{j1} + \dots + \beta_k X_{jk}}} \quad (14)$$

řeklinde ifade edilir (11, 12, 20, 27). Burada;

n: birim sayısını,

j: 1, 2, ..., n

$P(Y_j=1)$: j. birimin incelenen kategoriye eřit olma olasılıęını ya da incelenen olay ile ilgili pozitif cevap verme olasılıęını,

β_0 : Baęımsız deęiřkenler sıfır deęerini aldıęında baęımlı deęiřkenin deęerini, yani sabiti,

$\beta_1, \beta_2, \dots, \beta_k$: Baęımsız deęiřkenlerin regresyon katsayılarını,

$X_{j1}, X_{j2}, \dots, X_{jk}$: j. birime ait baęımsız deęiřkenleri,

k: Baęımsız deęiřken sayısını,

e = 2,718281828 sayısını gstermektedir (35, 38, 43, 51).

2.4. Sıralı Lojistik Regresyon Modeli (Ordinal Logistic Regression)

Baęımlı deęiřkenin sıralı lekli olduęu durumlarda uygulanan bir yntemdir. Sıralı lekli baęımlı deęiřken en az  kategoride gzlenen deęerler iermelidir; hafif, orta, aęır; kt, iyi, ok iyi gibi. Sıralı lekli veriler kodlanırken kategorilerin, deęiřkenin doęal sıralama yapısına gre kodlama

yapılır. Bu isimsel kategorilerin kod değerleri de aynı büyüklük yapısını izlemelidir (20, 27, 34, 38, 43).

2.5. İsimsel Lojistik Regresyon Modeli (Nominal Logistic Regression)

Bağımlı değişkenin isimsel ölçekli olduğu ve en az üç kategoride gözlemlendiği durumlarda uygulanır. Bu modelde bağımlı değişkenin kategorileri arasında doğal bir sıralama yoktur ve kodlama yaparken kodlamanın bir doğal sıralama içinde olması şart değildir (43).

Sıralı ve isimsel lojistik regresyon modelleri ikili lojistik regresyon modelinin genişletilmiş şeklidir. Sıralı ve isimsel lojistik regresyon modellerine çoklu lojistik regresyon modelleri de denir. Bu modellerde bağımlı değişken iki kategoriden fazla kategoriye sahip olduğundan ilk veya son kategori referans kategori olarak seçilir ve diğer kategorilerle arasındaki ilişki olasılık olarak incelenir. Çoklu lojistik regresyon modelinde, bağımlı değişkenin kategorileri h ve referans kategorisi h_0 ile gösterilir ise, çoklu lojistik regresyon modeli;

$$P(Y_j = h) = \frac{e^{\beta_{h0} + \beta_{h1}X_{j1} + \beta_{h2}X_{j2} + \dots + \beta_{hk}X_{jk}}}{1 + \sum_{h=1}^{M-1} e^{\beta_{h0} + \beta_{h1}X_{j1} + \beta_{h2}X_{j2} + \dots + \beta_{hk}X_{jk}}} \quad (15)$$

şeklinde ifade edilir (20, 21, 29, 38, 45).

Referans kategori içinde çoklu lojistik regresyon modeli ise;

$$P(Y_j = h_0) = \frac{1}{1 + \sum_{h=1}^{M-1} e^{\beta_{h0} + \beta_{h1}X_{j1} + \beta_{h2}X_{j2} + \dots + \beta_{hk}X_{jk}}} \quad (16)$$

şeklinde ifade edilir.

Burada;

n: birim sayısını,

j: 1, 2, ..., n

M: bağımlı değişkenin sahip olduğu kategori sayısını,

h=1, 2, ..., M-1

h_0 : bağımlı değişkende kullanılan referans kategoriyi, (ya ilk ya da son kategori olarak alınır, $h_0=0$, $h_0=M$).

$P(Y_j=h)$: j. birimin incelenen kategoriye eşit olma olasılığını,

$P(Y_j=h_0)$: j. birimin referans kategorisine eşit olma olasılığını,

β_{h_0} : h. kategori için bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini, yani sabiti,

$\beta_{h_1}, \beta_{h_2}, \dots, \beta_{h_k}$: h. kategori için bağımsız değişkenlerin regresyon katsayılarını,

$X_{j1}, X_{j2}, \dots, X_{jk}$: j. birime ait bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

e = 2,718281828 sayısını göstermektedir.

Referans kategori için;

$$e^{\beta_{h_0 0} + \beta_{h_0 1} X_1 + \beta_{h_0 2} X_2 + \dots + \beta_{h_0 k} X_k} = 1 \quad (17)$$

olarak alınır (38).

2.6. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modelleri

Sınırlandırılmış ve koşullu lojistik regresyon analizinde kullanılan bağımlı değişken, veri yapısı ve modeller aşağıda açıklanmıştır.

Sınırlandırılmış İkili Kategorik Veri (Truncated Binary Data): Birden fazla gruptan oluşan, her grupta iki ya da daha fazla birimin olduğu ve her grup için, incelenen konu hakkında pozitif cevabın elde edilebildiği bir veri yapısıdır (22, 26, 33, 42).

Sınırlandırılmış ve koşullu lojistik regresyon analizlerinde bağımlı değişken olarak, sınırlandırılmış ikili kategorik veri kullanılır.

2.6.1. Koşullu Lojistik Regresyon Modeli (Conditional Logistic Regression)

Koşullu lojistik regresyon modeli;

$$P(Y_{ij} = 1) = \frac{e^{a_i + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \dots + \beta_k X_{ijk}}}{1 + e^{a_i + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \dots + \beta_k X_{ijk}}} \quad (18)$$

şeklinde ifade edilir (9, 14, 15, 30, 53).

Burada;

$i=1,2, \dots, G$

G: grup sayısını,

$j=1, 2, \dots, n_i$

n_i : i. gruptaki birim sayısını,

$P(Y_{ij}=1)$: i. gruptaki j. birimin pozitif cevap gösterme olasılığını,

α_i : i. grubun etkisini,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_{ij1}, X_{ij2}, \dots, X_{ijk}$: i. gruptaki j. birime ait bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

e = 2,718281828 sayısını göstermektedir (6, 13, 36, 50).

2.6.2. Sınırlandırılmış Lojistik Regresyon Modeli

(Truncated Logistic Regression)

Sınırlandırılmış lojistik regresyon modeli;

$$P(Y_{ij} = 1) = \frac{e^{\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \dots + \beta_k X_{ijk}}}{1 + e^{\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \dots + \beta_k X_{ijk}}} \quad (19)$$

şeklinde ifade edilir (42).

Burada;

$i=1,2, \dots, G$

G: grup sayısını,

$j=1, 2, \dots, n_i$

n_i : i. gruptaki birim sayısını,

$P(Y_{ij}=1)$: i. gruptaki j. birimin pozitif cevap gösterme olasılığını,

β_0 : Bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini, yani sabiti,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_{ij1}, X_{ij2}, \dots, X_{ijk}$: i. gruptaki j. birime ait bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

e = 2,718281828 sayısını göstermektedir.

Sınırlandırılmış lojistik regresyon modeli, her grup için en az bir pozitif cevabın olmasını şart koşturmaktadır (12, 26, 42).

2.7. Parametrelerin Önemliliğinin ve Modelin Uygunluğunun Test Edilmesi

Tahmin edilen β katsayılarının önemliliğinin test edilmesi $H_0: \beta=0$ hipotezinin test edilmesidir. Bu amaçla ileri sürülen testler;

1. Wald testi

2. Benzerlik Oranı (Likelihood Ratio) testidir.

Bu testler asimptotik olarak aynı sonucu verirler. Örnek hacmi büyüdükçe testlerin sonuçları çok yakın çıkmaktadır (5, 20, 38).

2.7.1. Wald Testi

Wald testi en büyük benzerlik tahminlerinin (Maximum Likelihood Estimations) asimptotik olarak normal dağılım gösterdiği varsayımına dayanır. Regresyon katsayılarının, standart hatasına oranı;

$$z = \frac{\beta}{SE_{\beta}} \quad (41)$$

Wald istatistiği olarak adlandırılır. Bu durumda Wald istatistiği Standart Normal Dağılım (SND) gösterir ve SND'nin kritik değerleri ile karşılaştırılarak önemliliği belirlenir. Aynı zamanda Wald istatistiği;

$$w = z^2 = \left(\frac{\beta}{SE_{\beta}} \right)^2 \quad (42)$$

olarak da kullanılabilir. Bu durumda ise Wald test istatistiği 1 serbestlik dereceli kıkare dağılımı gösterir ve 1 serbestlik dereceli kıkare dağılımının kritik değerleri ile karşılaştırılarak önemliliği belirlenir (1, 5).

2.7.2. Benzerlik Oranı Testi

X_1, X_2, \dots, X_k tane bağımsız değişken içeren bir regresyon analizinde, oluşturulan iki regresyon modelinin birinci modelinde X_1, X_2, \dots, X_v tane bağımsız değişken, ikinci modelinde ise $X_1, X_2, \dots, X_v, X_{v+1}, X_{v+2}, \dots, X_{v+m}$ ($v+m=k$) olmak üzere tüm bağımsız değişkenler olsun. Bu iki modelde de bağımsız değişkenlerin katsayılarını birinci model için β_v vektörü ikinci model içinde β_{v+m} vektörü gösterebilir. $X_{v+1}, X_{v+2}, \dots, X_{v+m}$ bağımsız değişkenlerinin katsayılarını sıfıra eşit olup olmadığını eşanlı (simültane) olarak test eden benzerlik oran test istatistiği;

$$LR = -2 \left(\log \left[\frac{L(\beta_v)}{L(\beta_{v+m})} \right] \right) = -2(\log[L(\beta_v)] - \log[L(\beta_{v+m})]) \quad (43)$$

şeklinde ifade edilir.

Burada;

$L(\beta_v)$: v tane bağımsız değişken içeren birinci model için en büyük benzerlik fonksiyonunu,

$L(\beta_{v+m})$: $v+m$ tane bağımsız değişken içeren ikinci model için en büyük benzerlik fonksiyonunu göstermektedir.

Benzerlik oran istatistiği LR, $(v + m) - v = m$ serbestlik dereceli ki-kare dağılımı gösterir ve m serbestlik dereceli ki-kare dağılımının kritik değerleri ile karşılaştırılarak β_{v+m} vektörünün son m tane katsayısının sıfıra eşit olup olmadığı test edilir.

Verilere uygulanan modelin uygunluğunu test etmek için model parametrelerinin tümünün sıfır olduğu hipotezi test edilir. Burada uygulanan test benzerlik oranı testidir. Birinci model olarak sabit dışında (β_0) hiçbir parametre modele katılmaz. İkinci model olarak ise tüm parametreler modele katılır. Birinci ve ikinci model için hesaplanan en büyük benzerlik fonksiyonları elde edilerek benzerlik oranı LR test istatistiği yukarıdaki gibi hesaplanır. Uygunluk testi için hesaplanan benzerlik oranı test istatistiği toplam parametre sayısının 1 eksiği kadar ya da modele katılan bağımsız değişken sayısı kadar serbestlik dereceli kikare dağılımı gösterir ve kikare dağılımının kritik değerleri ile karşılaştırılarak modelin uygunluğu test edilir (20, 38, 49).

3. GEREÇ ve YÖNTEM

Bu çalışmada, bölünmüş ikili veri yapısına uygulanan koşullu ve sınırlandırılmış lojistik regresyon yöntemlerinin karşılaştırılması amaçlandığından, bu bölümde bu iki yöntem incelenmiş, yöntemler ayrıntıları ile açıklanmıştır. Uygulamada iki modelin benzerliklerini ve farklılıklarını göstermek amacıyla da Eskişehir ili Trafik İl Şube Müdürlüğü'nden özel izinle elde edilmiş trafik kazaları verileri kullanılmıştır. Bu veri yapısı sadece önden çarpma sonucu oluşan ve ölümlerle sonuçlanan otomobil kazalarını içermektedir. Koşullu ve sınırlandırılmış lojistik regresyon yöntemleri ve uygulamada kullanılan veri yapısı ayrıntılı olarak aşağıdaki alt bölümlerde açıklanmıştır.

3.1. Veri Yapısı

Uygulamada kullanılan veri yapısı; düzeltilmemiş, sadece kaza anındaki raporlardan elde edilmiş önden çarpma sonucu oluşan ve ölümlerle sonuçlanan otomobil kazalarını içermektedir.

Bu veri yapısında;

-Her bir kaza birbirinden bağımsızdır.

-Grup olarak kazaya karışan otomobiller ele alınmıştır.

-Grup sayısı 20'dir. Başka bir anlatımla önden çarpma sonucu meydana gelen, çeşitli kazalara karışan toplam 20 otomobil 20 ayrı grup olarak ele alınmıştır.

-Her bir otomobilde (her bir grupta) en az iki birey yer almış ve en az bir ölüm meydana gelmiştir. 20 otomobilde toplam 76 birey yer almaktadır.

-Kazalarda incelenen olay ya da pozitif cevap, ölüm olarak belirlenmiş ve $Y_{ij}=1$ i. kazadaki j. bireyin öldüğünü, $Y_{ij}=0$ ise i. kazada j. bireyin yaşadığını göstermektedir.

Kazalarda meydana gelen ölümler bağımlı değişken olarak, ölümleri etkileyen ve ölüm riskini değiştirebilen faktörler bağımsız değişken olarak aşağıda açıklanmıştır. Uygulamada, koşullu ve sınırlandırılmış lojistik regresyon modelleri kullanılarak faktörlerin ölüm riski üzerine etkileri araştırılmış, iki yöntem arasındaki benzerlikler ve farklılıklar uygulama üzerinde gösterilerek hangi yöntemin hangi durumda kullanılacağı, hangisinin daha üstün olduğu belirtilmiştir.

Bağımlı değişken;

$$Y_{ij} = \begin{cases} 0 & \text{i. kazadaki j. birey yaşıyor ise} \\ 1 & \text{i. kazadaki j. birey ölmüş ise} \end{cases}$$

Bağımsız değişkenler;

1- Emniyet kemeri takmak;

$$X_{ij1} = \begin{cases} 0 & \text{i. kazadaki j. birey emniyet kemeri takmış ise} \\ 1 & \text{i. kazadaki j. birey emniyet kemeri takmamış ise} \end{cases}$$

2- Kazaya karışan bireyin yaşı;

$$X_{ij2} = \{i. kazadaki j. bireyin yaşı\}$$

3- Kazaya karışan bireyin otomobildeki oturma yeri

$$X_{ij3} = \begin{cases} 0 & \text{i. kazadaki j. birey arkada oturuyor ise} \\ 1 & \text{i. kazadaki j. birey önde oturuyor ise} \end{cases}$$

4- Kazaya karışan otomobilin çarpışma hızı

$$X_{ij4} = \{i. kazadaki aracın çarpma hızı\}$$

Burada, Y_{ij} sınırlandırılmış ikili bağımlı değişkendir. X_{ij1} , X_{ij2} ve X_{ij3} birey düzeyinde bağımsız değişkenlerdir. X_{ij4} ise grup düzeyinde (otomobil düzeyinde) bağımsız değişken olduğundan i. otomobildeki tüm bireyler için aynı değeri içerir ve bireyden bireye değişmez.

3.2. Koşullu ve Sınırlandırılmış Lojistik Regresyon Analizlerinde Kullanılan Bağımlı ve Bağımsız Değişkenler

Koşullu ve sınırlandırılmış lojistik regresyon analizinde kullanılan bağımlı değişken sınırlandırılmış ikili cevap içeren kategorik değişkendir. i. gruptaki j. birey için incelenen olay ya da riskin gözlenmesi $Y_{ij}=1$ ile diğer durumlar ise $Y_{ij}=0$ ile gösterilir (8, 12, 18, 42, 46).

Bağımsız değişkenler ise grup düzeyinde bağımsız değişkenler ve birey düzeyinde bağımsız değişkenler olmak üzere iki ana gruba ayrılırlar.

Grup Düzeyindeki Bağımsız Değişkenler: Herhangi bir grupta tüm bireyler için aynı değere sahip (grup için sabit değer) ve grup düzeyinde bilgi içeren bağımsız değişkenlerdir (42). Uygulamada, kazaya karışan otomobilin çarpma hızı otomobil düzeyinde (grup düzeyinde) bağımsız değişken olarak alınmıştır.

Birey Düzeyindeki Bağımsız Değişkenler: Bireyden bireye farklı değerler alan, grup düzeyinde bilgi içermeyen bağımsız değişkenlerdir (12, 42). Uygulamada, kazaya karışan bireylerin emniyet kemeri takma durumu, yaşı ve otomobil içerisindeki oturma yeri birey düzeyinde bağımsız değişken olarak alınmıştır.

Koşullu lojistik regresyon analizi, modele katılan tüm bağımsız değişkenlerin, birey düzeyindeki bağımsız değişkenler olduğunu varsayar (12, 37). Uygulamada, koşullu lojistik regresyon analizinde kullanılan bağımsız değişkenler; kazaya karışan bireylerin emniyet kemeri takma durumu, yaşı ve otomobildeki oturma yeridir.

Sınırlandırılmış lojistik regresyon analizinde, modele katılan bağımsız değişkenler hem grup düzeyinde hem de birey düzeyinde bağımsız değişkenler olabilir (42). Uygulamada kazaya karışan bireylerin emniyet kemeri takma durumu, yaşı ve otomobildeki oturma yeri ile birlikte otomobilin çarpma hızı sınırlandırılmış lojistik regresyon modelinde analiz edilmiştir.

3.3. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerinde Regresyon Katsayıların Tahminlenmesi

Sınırlandırılmış ve koşullu lojistik regresyon modellerindeki regresyon katsayılarının tahminlenmesinde en büyük benzerlik fonksiyonu (maximum likelihood function) kullanılmıştır. Regresyon katsayılarının değerleri en büyük benzerlik tahmin değerleri olarak hesaplanmıştır (3, 39, 48, 49, 52).

3.3.1. Koşullu Lojistik Regresyon Modelinde Regresyon Katsayılarının Tahminlenmesi

Koşullu lojistik regresyon modeli ve parametre tahminleri aşağıdaki eşitliklerle hesaplanmıştır.

$$P(Y_{ij} = 1) = \frac{e^{\alpha_i + \beta'X_{ij}}}{1 + e^{\alpha_i + \beta'X_{ij}}} \quad (20)$$

$$X_{ij} = \begin{bmatrix} X_{ij1} \\ X_{ij2} \\ \cdot \\ \cdot \\ \cdot \\ X_{ijk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}. \quad (21)$$

Burada;

$i=1,2, \dots, G$

G : grup sayısını,

$j=1, 2, \dots, n_i$

n_i : i . gruptaki birey sayısını,

$P(Y_{ij}=1)$: i . gruptaki j . bireyin pozitif cevap gösterme olasılığını,

α_i : i . grubun etkisini,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_{ij1}, X_{ij2}, \dots, X_{ijk}$: i. gruptaki j. bireye ait bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

e = 2,718281828 sayısını göstermektedir.

Koşullu lojistik regresyon modeli nuisance parametreler olan α_i grup etkilerini, yeterli (sufficient) istatistikler olan toplam pozitif gözlem sayılarına koşullanarak en büyük benzerlik fonksiyonundan elimine eder. Yeterli istatistikler olan toplam pozitif gözlem sayıları;

$$r_i = \sum_{j=1}^{n_i} Y_{ij}, \quad i=1, 2, \dots, N \quad (22)$$

şeklinde ifade edilir.

Buradan koşullu referans seti;

$$R_{r_i} = \left\{ \mathbf{u} = [u_1, u_2, \dots, u_{n_i}]' : u_m \in \{0, 1\}, \sum_{m=1}^{n_i} u_m = r_i \right\} \quad (23)$$

şeklinde ifade edilir. Γ_{r_i} i. grup için toplamları r_i olan n_i tane ikili cevap içeren bağımlı değişkenin olası tüm vektörlerinin setini gösterir. Buradan koşullu lojistik regresyon modelinde regresyon katsayılarının hesaplanması için en büyük benzerlik fonksiyonu;

$$L(\boldsymbol{\beta}) = \prod_{i=1}^G \frac{\prod_{j=1}^{n_i} [\exp(\boldsymbol{\beta}' \mathbf{X}_{ij})]^{Y_{ij}}}{\sum_{\mathbf{u} \in R_{r_i}} \left[\prod_{m=1}^{n_i} [\exp(\boldsymbol{\beta}' \mathbf{X}_{im})]^{u_m} \right]} \quad (24)$$

şeklinde ifade edilir. Bu fonksiyonu en büyükleyen β değerleri Newton-Raphson algoritması ile elde edilmektedir.

Uygulamada ve analizlerde, koşullu lojistik regresyon analizinin parametre tahminleri, EGRET for WINDOWS Version 2.0.3 paket programı kullanılarak elde edilmiştir (4, 41, 44, 47).

3.3.2. Sınırlandırılmış Lojistik Regresyon Modelinde Regresyon Katsayılarının Tahminlenmesi

Sınırlandırılmış lojistik regresyon modeli ve parametre tahminleri aşağıdaki eşitliklerle hesaplanmıştır.

$$P(Y_{ij} = 1) = \frac{e^{\beta'X_{ij}}}{1 + e^{\beta'X_{ij}}} \quad (25)$$

Burada;

$$X_{ij} = \begin{bmatrix} 1 \\ X_{ij1} \\ X_{ij2} \\ \cdot \\ \cdot \\ \cdot \\ X_{ijk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad (26)$$

$i=1,2, \dots, G$

G : grup sayısını,

$j=1, 2, \dots, n_i$

n_i : i . gruptaki birey sayısını,

$P(Y_{ij}=1)$: i. gruptaki j. bireyin pozitif cevap gösterme olasılığını,

β_0 : Bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini,

$\beta_1, \beta_2, \dots, \beta_k$: Bağımsız değişkenlerin regresyon katsayılarını,

$X_{ij1}, X_{ij2}, \dots, X_{ijk}$: i. gruptaki j. bireye ait bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

e = 2,718281828 sayısını göstermektedir (4, 10, 24, 42).

Sınırlandırılmış lojistik regresyon modeli;

$$P(Y_{ij} = 1) = \frac{e^{\beta'X_{ij}}}{1 + e^{\beta'X_{ij}}} = p(\beta, X_{ij}) = 1 - q(\beta, X_{ij}) \quad (27)$$

şeklinde ifade edilir ise, sınırlandırılmış lojistik regresyon modelinde regresyon katsayılarının hesaplanması için en büyük benzerlik fonksiyonu;

$$L(\beta) = \prod_{i=1}^G \frac{\prod_{j=1}^{n_i} p(\beta, X_{ij})^{Y_{ij}} q(\beta, X_{ij})^{1-Y_{ij}}}{1 - \prod_{j=1}^{n_i} q(\beta, X_{ij})} \quad (28)$$

şeklinde ifade edilir.

Sınırlandırılmış lojistik regresyon modeli her grup için en az bir pozitif cevabın gözlenmesi koşulunu öngördüğü için, konvansiyonel lojistik regresyon modelindeki en büyük benzerlik fonksiyonuna

$$1 - \prod_{j=1}^{n_i} q(\beta, \mathbf{X}_{ij}) \quad (29)$$

böleni eklenerek sınırlandırılmış lojistik regresyon modeli için yukarıdaki en büyük benzerlik fonksiyonu elde edilir (42, 47).

Sınırlandırılmış lojistik regresyon yönteminin parametre tahminlerini hesaplayan istatistiksel paket programına literatürlerde rastlanılmamıştır. Bu nedenle, bu çalışmada sınırlandırılmış lojistik regresyon analizini yapan program, MATLAB 5.3 programlama dili kullanılarak yazılmıştır. Bu programda, skor fonksiyon ve Fisher information matris elde edilerek Newton-Raphson algoritması ile sınırlandırılmış lojistik regresyon yönteminin parametre tahminleri hesaplanmıştır (39, 41,48, 49).

Skor ve information matrislerinin hesaplanmasında $P_i(\beta)$ ve $\mu_i(\beta)$ eşitlikleri aşağıdaki gibi belirlenir.

$$P_i(\beta) = 1 - Q_i(\beta) = 1 - \prod_{j=1}^{n_i} q(\beta, \mathbf{X}_{ij}) \quad (30)$$

şeklinde ifade edilir ve

$$\mu_i(\beta) = P_i(\beta)^{-1} \sum_{j=1}^{n_i} p(\beta, \mathbf{X}_{ij}) \mathbf{X}_{ij} \quad (31)$$

şeklinde ifade edilir ise sınırlandırılmış lojistik regresyon yöntemi için skor fonksiyon;

$$S(\beta) = \frac{\partial \log\{L(\beta)\}}{\partial \beta} = \sum_{i=1}^G \left\{ \sum_{j=1}^{n_i} \mathbf{X}_{ij} Y_{ij} - \mu_i(\beta) \right\} \quad (32)$$

şeklinde ifade edilir. Skor fonksiyon; sınırlandırılmış lojistik regresyon yönteminin en büyük benzerlik fonksiyonunun logaritmasının parametrelere göre türevi alınmış şeklidir;

$$S(\hat{\beta}) = 0 \quad (33)$$

eşitliğini sağlayan $\hat{\beta}$ değerleri, β parametrelerinin en büyük benzerlik tahminleridir.

Aynı zamanda yukarıdaki eşitlikler kullanılarak Fisher information matrisi;

$$I(\beta) = \sum_{i=1}^G \left\{ P_i(\beta)^{-1} \sum_{j=1}^{n_i} p(\beta, X_{ij}) q(\beta, X_{ij}) X_{ij} X_{ij}' - Q_i(\beta) \mu_i(\beta) \mu_i(\beta)' \right\} \quad (34)$$

şeklinde elde edilir (42).

En büyük benzerlik tahminleri olan $\hat{\beta}$ değerlerinin varyans kovaryans matrisi, Fisher information matrisinin tersidir ve

$$V(\hat{\beta}) = I(\hat{\beta})^{-1} \quad (35)$$

şeklinde ifade edilir (42).

MATLAB 5.3 programlama dili kullanılarak yazılan programda parametre tahminleri için;

$$\hat{\beta}_{t+1} = \hat{\beta}_t + [I(\hat{\beta}_t)^{-1} S(\hat{\beta}_t)] \quad (36)$$

şeklinde yinelenen Newton-Raphson iterasyonları kullanılarak parametre tahminleri elde edilmiştir (2, 19, 48).

Burada;

$\hat{\beta}_{t+1}$: (t+1). iterasyondaki parametre tahminleri,

$\hat{\beta}_t$: t. iterasyondaki parametre tahminleri,

$I(\hat{\beta}_t)^{-1}$: t. iterasyonda elde edilmiş tahmin değerlerinin varyans kovaryans matrisi,

$S(\hat{\beta}_t)$: t. iterasyonda elde edilmiş tahmin değerlerinin skor fonksiyonu değerleridir.

İlk iterasyon için;

$$\hat{\beta}_1 = \hat{\beta}_0 + [I(\hat{\beta}_0)^{-1}S(\hat{\beta}_0)] \quad (37)$$

şeklinde ifade edilir.

Burada;

$$\hat{\beta}_0 = \mathbf{0} \quad (38)$$

sıfır matrisi olarak alınır. Başka bir anlatımla iterasyona başlarken tüm parametrelere başlangıç değeri olarak sıfır değeri atanmıştır (19).

İterasyon sayısının ne olacağına yada kaçınıcı iterasyonda duracağımıza karar verebilmek için iki tip tolerans düzeyi kullanılmıştır (19).

Birinci tolerans düzeyi;

$$\max \left| \hat{\beta}_{m,t+1} - \hat{\beta}_{m,t} \right| < \varepsilon \quad (39)$$

şeklinde ifade edilir.

Burada;

$m=0, 1, 2, \dots, k$

k : bağımsız değişken sayısını,

$\hat{\beta}_{m,t+1}$: m . bağımsız değişkenin $(t+1)$. iterasyondaki tahmin değerini,

$\hat{\beta}_{m,t}$: m . bağımsız değişkenin t . iterasyondaki tahmin değerini,

ε : tolerans seviyesini göstermektedir.

Tolerans seviyesi 0.0001'e eşit ya da daha küçük bir sayıdır. Sonuçta, $(t+1)$. iterasyonda hesaplanan tahmin değerleri ile t . iterasyonda hesaplanan tahmin değerleri arasındaki mutlak farkın 0.0001'den daha küçük ise iterasyon durdurulmuştur ve $(t+1)$. iterasyondaki tahmin değerleri kullanılmıştır (19).

İkinci tolerans düzeyi;

$$S(\hat{\beta}_{m,t+1}) < \varepsilon \quad (40)$$

şeklinde ifade edilir.

Burada;

$S(\hat{\beta}_{m,t+1})$: m. bağımsız değişkenin (t+1). iterasyondaki tahmin değerinin skor fonksiyonu değerini,

ϵ : tolerans seviyesini göstermektedir.

Tolerans seviyesi 0.0001'e eşit ya da daha küçük bir sayıdır. Sonuçta; (t+1). iterasyonda hesaplanan tahmin değerlerinin skor fonksiyon değerleri 0.0001'den daha küçük olduğunda iterasyon durdurulur ve (t+1). iterasyon sonundaki tahmin değerleri kullanılır (19). Parametre tahminleri için programda iki tolerans düzeyi de kullanılmıştır.

Parametre tahminlerinin anlamlılığı ve modelin uygulununun test edilmesi için Wald testinden yararlanılmıştır.

4. BULGULAR

Bu arařtırmada, kořullu ve sınırlandırılmıř lojistik regresyon modeli ile analiz edilen 2000-2001 yıllarına ait önden çarpma sonucu oluřan otomobil kazalarına ait veriler Çizelge 4.1.'de verilmiřtir.

Çizelge 4.1. Eskiřehir 2000-2001 Yılı Önden Çarpma Sonucu Oluřan Otomobil Kazaları Verileri

Otomobil	Bireyin Durumu (Yij)	Bireyin Emniyet Kemerini Takma Durumu (Xij1)	Bireyin Yaşı (Xij2)	Bireyin Otomobildeki Yeri (Xij3)	Otomobilin Çarpma Hızı (Xij4)
1	1	1	43	1	110
	0	0	31	0	110
	0	0	26	0	110
2	1	1	24	1	140
	1	1	27	0	140
	1	1	18	0	140
	0	1	32	1	140
3	1	0	39	1	145
	1	0	35	1	145
	1	1	28	0	125
	0	1	42	0	125
	0	1	16	0	125
4	1	1	27	1	80
	0	0	38	0	80
	0	1	46	0	80
5	1	1	54	1	110
	1	1	33	0	110
	0	1	48	0	110
	0	1	31	0	110
6	1	0	21	1	130
	1	1	24	1	130
	0	1	19	0	130
7	0	0	43	1	100
	1	1	45	0	100
8	1	1	22	1	140
	0	0	24	1	140
	1	1	18	0	140
	1	1	26	0	140
	0	1	45	0	140
9	0	0	21	1	100
	1	1	45	1	100
	1	1	56	0	100

Otomobil	Bireyin Durumu (Yij)	Bireyin Emniyet Kemerı Takma Durumu (Xij1)	Bireyin Yaşı (Xij2)	Bireyin Otomobildeki Yeri (Xij3)	Otomobilin Çarpma Hızı (Xij4)
10	1	1	42	1	70
	0	0	48	1	70
	0	1	36	0	70
	0	1	29	0	70
	0	1	33	0	70
11	1	0	26	1	95
	0	0	28	1	95
	1	1	32	0	95
	0	1	15	0	95
12	1	0	51	1	130
	1	0	49	1	130
	0	1	57	0	130
	0	1	41	0	130
13	1	1	36	1	140
	1	1	41	1	140
	1	1	27	0	140
	0	1	19	0	140
	0	1	23	0	140
14	0	0	26	1	85
	1	1	24	1	85
	0	1	28	0	85
15	0	1	56	1	95
	0	0	62	1	95
	1	1	54	0	95
	0	1	54	0	95
16	1	1	25	1	145
	1	1	26	1	145
	0	1	24	0	145
	1	1	24	0	145
	0	1	28	0	145
17	1	1	27	1	130
	1	1	28	1	130
18	0	0	25	1	120
	1	0	24	1	120
	1	1	26	0	120
19	0	0	26	1	150
	1	0	28	1	150
	1	0	29	0	150
	1	1	32	0	150
	1	1	35	0	150
20	0	0	56	1	100
	1	1	54	1	100
	0	1	44	0	100
	1	1	47	0	100

Uygulamada, önden çarpma sonucu oluşan otomobil kazalarında ölüm riski üzerine etkileri araştırılan sabit ve bağımsız değişkenler modellere katılma durumlarına göre Çizelge 4.2.'de verilmiştir.

Çizelge 4.2. Uygulamada Kullanılan Sabit ve Bağımsız Değişkenlerin Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerine Katılma Durumu

Değişkenler	Koşullu Lojistik Regresyon Modeli	Sınırlandırılmış Lojistik Regresyon Modeli
Sabit	YOK	VAR
Bireyin Emniyet Kemerini Takma Durumu	VAR	VAR
Bireyin Yaşı	VAR	VAR
Bireyin Otomobildeki Oturma Yeri	VAR	VAR
Otomobilin Çarpma Hızı	YOK	VAR

4.1. Koşullu Lojistik Regresyon Modeli Bulguları

Koşullu lojistik regresyon modeli kullanılarak, elde edilen trafik kazalarına ait verilerin analiz sonuçları Çizelge 4.3.'de verilmiştir.

Bireyin Emniyet Kemerini Takma Durumu, Bireyin Yaşı ve Bireyin Otomobildeki Oturma Yeri bağımsız değişken olarak ele alınmış ve bu değişkenlerin ölüm riski üzerine etkileri koşullu lojistik regresyon modelinde analiz edilmiştir. **Sabit**, koşullu lojistik regresyon modelinde parametre tahminleri yapılırken en büyük benzerlik fonksiyonundan elimine edildiğinden ve **Otomobilin Çarpma Hızı** grup düzeyinde (otomobil düzeyinde) bağımsız değişken olduğundan bu modelde kullanılmamıştır.

Çizelge 4.3. Koşullu Lojistik Regresyon Modeli Analiz Sonuçları

DEĞİŞKEN	KATSAYILAR	STD. HATA	WALD İSTATİSTİĞİ	P	ODDS ORANI	%95 GÜVEN ARALIĞI (CI)	
Bireyin Emniyet Kemerini Takma Durumu	1.8615	0.7942	5.4937	0.0191	6.4336	1.3565	30.5129
Bireyin Yaşı	0.0184	0.0332	0.3079	0.5790	1.0186	0.9544	1.0871
Bireyin Otomobildeki Oturma Yeri	1.6342	0.6753	5.8562	0.0155	5.1252	1.3643	19.2540

Çizelge 4.3.'de görüldüğü gibi;

Kazaya karışan bireylerin emniyet kemeri takmamış olmasının ($\hat{\beta}_1 = 1.8615 \pm 0.7942$) önemli düzeyde ölüm riski üzerine etkisi görülmektedir. Emniyet kemeri takmayan bireylerin, emniyet kemeri takan bireylere göre önden çarpma ile oluşan bir kazada ölüm riskinin 6.4336 kat (ODDS oranı= 6.4336, CI=[1.3565, 30.5129]) daha fazla olduğu görülmektedir (p=0.0191*).

Kazaya karışan bireylerin yaşının ($\hat{\beta}_2 = 0.0184 \pm 0.0332$) ölüm riski üzerine etkisi yoktur (ODDS oranı= 1.0186, CI=[0.9544, 1.0871]) (p=0.5790 n.s.).

Bireyin otomobildeki oturma yerinin ($\hat{\beta}_3 = 1.6342 \pm 0.6753$) önemli düzeyde ölüm riski üzerine etkisi görülmektedir. Otomobilin önünde oturan bireylerin arkada oturan bireylere göre, önden çarpışma ile oluşan bir kazada ölüm riskinin 5.1252 kat (ODDS oranı= 5.1252 CI=[1.3643, 19.2540]) daha fazla olduğu bulunmuştur (p=0.0155*).

4.2. Sınırlandırılmış Lojistik Regresyon Modeli Bulguları

Sınırlandırılmış lojistik regresyon modeli kullanılarak, trafik kazalarına ait verilerin analiz sonuçları Çizelge 4.4.'de verilmiştir.

Bireyin Emniyet Kemerini Takma Durumu, Bireyin Yaşı, Bireyin Otomobildeki Oturma Yeri ve Otomobilin Çarpma Hızı bağımsız değişken

olarak ele alınmış ve bu değişkenlerin ölüm riski üzerine etkileri sınırlandırılmış lojistik regresyon modelinde analiz edilmiştir.

Çizelge 4.4. Sınırlandırılmış Lojistik Regresyon Modeli Analiz Sonuçları

DEĞİŞKEN	KATSAYILAR	STD. HATA	WALD İSTATİSTİĞİ	P	ODDS ORANI	%95 GÜVEN ARALIĞI (CI)	
Sabit	-7.7144	2.6483	8.4853	0.0036	-	-	-
Bireyin Emniyet Kemerini Takma Durumu	1.8315	0,7433	6.0705	0.0137	6.2435	1.4542	26.8028
Bireyin Yaşı	0.0091	0.0264	0.1193	0.7298	1.0091	0.9581	1.0628
Bireyin Otomobilindeki Oturma Yeri	1.8195	0.6813	7.1304	0.0076	6.1690	1.6224	23.4538
Otomobilin Çarpma Hızı	0.0436	0.0173	7.1768	0.0074	1.0446	1.0096	1.0806

Çizelge 4.4.'de görüldüğü gibi sınırlandırılmış lojistik regresyon modelinde;

Sınırlandırılmış lojistik regresyon modelinde koşullu lojistik regresyon modelinden farklı olarak sabit katsayısı yer almaktadır. Bu katsayı ($\hat{\beta}_0 = -7.7144 \pm 2.6483$) olarak hesaplanmıştır ($p=0.0036^{**}$). Diğer tüm bağımsız değişkenler sıfır değeri aldığı anda bu katsayı modele katılarak bireyin ölme riski hesaplanmaktadır.

Kazaya karışan bireylerin emniyet kemeri takmış olmasının ($\hat{\beta}_1 = 1.8315 \pm 0.7433$) önemli düzeyde ölüm riski üzerine etkisi görülmektedir. Emniyet kemeri takmayan bireylerin, emniyet kemeri takan bireylere göre önden çarpma ile oluşan bir kazada ölüm riskinin 6.2435 kat (ODDS oranı= 6.2435, CI=[1.4542, 26.8028]) daha fazla olduğu görülmektedir ($p=0.0137^*$).

Kazaya karışan bireylerin yaşının ($\hat{\beta}_2 = 0.0091 \pm 0.0264$) ölüm riski üzerine etkisi yoktur (ODDS oranı= 1.0091, CI=[0.9581, 1.0628]) ($p=0.7298^{n.s.}$).

Bireyin otomobildeki oturma yerinin ($\hat{\beta}_3 = 1.8195 \pm 0.6813$) önemli düzeyde ölüm riski üzerine etkisi görülmektedir. Otomobilin önünde oturan bireylerin arkada oturan bireylere göre, önden çarpma ile oluşan bir kazada ölüm riskinin 6.1690 kat (ODDS oranı= 6.1690 CI=[1.6224, 23,4538]) daha fazla olduğu bulunmuştur ($p=0.0076^{**}$).

Kazaya karışan otomobilin çarpma hızının ($\hat{\beta}_4 = 0.0436 \pm 0.0173$) önemli düzeyde ölüm riski üzerine etkisi görülmektedir. Otomobil hızı 1 km arttığında, önden çarpma ile oluşan bir kazada ölüm riskinin de 1.0446 kat (ODDS oranı= 6.2435 CI=[1.0096, 1.0806]) daha fazla olduğu bulunmuştur ($p=0.0074^{**}$). Otomobil hızı 20 km arttığında ölüm riski de;

$$e^{(20 \times 0.0436)} = e^{0.872} = 2.3916 \text{ kat artmaktadır.}$$

Çizelge 4.5.'de her iki modelden elde edilen parametre tahminleri verilmiştir.

Çizelge 4.5. Koşullu ve Sınırlandırılmış Lojistik Regresyon Modellerinden Elde Edilen Parametre Tahminlerinin Karşılaştırılması

DEĞİŞKEN	Koşullu Lojistik Regresyon Modeli	Sınırlandırılmış Lojistik Regresyon Modeli
Sabit	YOK	-7.7144
Bireyin Emniyet Kemerini Takma Durumu	1.8615	1.8315
Bireyin Yaşı	0.0184	0.0091
Bireyin Otomobildeki Oturma Yeri	1.6342	1.8195
Otomobilin Çarpma Hızı	YOK	0.0436

Çizelge 4.5.'de görüldüğü gibi;

Sabit ve otomobilin çarpma hızı koşullu lojistik regresyon modelinde yokken sınırlandırılmış lojistik regresyon modelinde analize katılmıştır.

5. TARTIŞMA ve SONUÇ

Sağlık alanında elde edilen verilerin analizinde aynı amaç için farklı modeller kullanılabilir. Veri yapısına ve amaca en uygun modelin seçiminde mevcut riskleri en aza indirgeyen modelin analizde kullanılması gerekmektedir (17).

Bağımlı değişken olarak sınırlandırılmış ikili veri yapısı kullanılan ve bu bağımlı değişken üzerine etkileri olan bağımsız değişkenlerin analizinde koşullu ve sınırlandırılmış lojistik regresyon modelleri kullanılmaktadır. Koşullu ve sınırlandırılmış lojistik regresyon modelleri aynı amaç için ve aynı tipte olan verilerin analizlerinde kullanılan iki farklı lojistik regresyon modelidir (42).

Bu çalışmada koşullu ve sınırlandırılmış lojistik regresyon modellerinin karşılaştırılması yapılmış ve bir uygulama üzerinde bu karşılaştırmalar gösterilmiştir.

Koşullu lojistik regresyon modelinde bağımlı değişken olarak sınırlandırılmış ikili veri kullanılmaktadır. Bu modelde sınırlandırılmış ikili cevap içeren bağımlı değişken ile bu değişkeni etkileyen bağımsız değişkenler (faktörler) arasındaki ilişki analiz edilmektedir. Modelde kullanılan bağımsız değişkenler mutlaka birey düzeyinde bağımsız değişkenler olmak zorundadır. Eğer grup düzeyindeki bağımsız değişkenlerin etkileri araştırılmak isteniyorsa ya da araştırmada grup düzeyinde bağımsız değişkenler varsa koşullu lojistik regresyon modeli kullanılamaz. Çünkü koşullu lojistik regresyon modelinde parametre tahminleri yapılırken, toplam pozitif cevap sayısına koşullanarak, grup etkileri olan sabit katsayısı ve grup düzeyindeki bağımsız değişkenler en büyük benzerlik fonksiyonundan elimine edilir (9, 15, 31, 37, 53).

Koşullu lojistik regresyon modelinde olduğu gibi, sınırlandırılmış lojistik regresyon modelinde de ikili cevap içeren bağımlı değişken ile bu değişkeni etkileyen bağımsız değişkenler (faktörler) arasındaki ilişki analiz edilir.

Bu modelde kullanılan bağımsız değişkenler üzerinde herhangi bir kısıtlama yoktur. Modeldeki bağımsız değişkenler hem grup hem de birey düzeyinde bağımsız değişkenler olabilir (7, 18, 28, 32, 33).

Koşullu ve sınırlandırılmış lojistik regresyon modellerinin kullanımının seçiminde;

-Araştırmada sadece birey düzeyinde bağımsız değişkenler varsa her iki modelden biri kullanılabilir (42).

-Eğer araştırmada birey düzeyinde bağımsız değişkenlerin yanı sıra grup düzeyinde bağımsız değişkenler varsa sınırlandırılmış lojistik regresyon modeli kullanılmalıdır (22, 32, 33).

Sınırlandırılmış lojistik regresyon modeli genelde tercih edilen modeldir. Çünkü hem grup hem de birey düzeyindeki bağımsız değişkenler analize katılabildiğinden, sınırlandırılmış lojistik regresyon modeli koşullu lojistik regresyon modeline göre, neden sonuç ilişkileri ile risk faktörlerinin etkilerinin belirlenmesinde daha etkin bir modeldir (42).

Uygulamada trafik kazalarına ait veriler her iki modelde de analiz edilmiş ve aşağıdaki sonuçlar elde edilmiştir.

Her iki modelde de;

-Emniyet kemeri takmanın ölüm riski üzerine önemli bir etkisinin olduğu bulunmuştur ($p < 0.05$). Emniyet kemeri takmamanın, O'Neill ve Barry de (1995) çalışmalarında belirttiği gibi, önden çarpma sonucu oluşan bir kazada ölüm riskini artırıcı bir etkiye sahip olduğu görülmüştür (42).

-Yapmış olduğumuz çalışmada, O'Neill ve Barry'nin de (1995) belirttiği gibi, birey yaşının, önden çarpma sonucu oluşan bir kazada ölüm riski üzerine etkisi olmadığı belirlenmiştir ($p > 0.05$) (42).

- Gomez (2000), O'Neill ve Barry (1995), Weiss'in (1992, 1993) yapmış oldukları çalışmalarla uyumlu olarak, bireyin arabadaki oturma yerinin ölüm riskini önemli düzeyde etkilediği ve önden çarpma sonucunda oluşan bir kazada ön koltukta oturma oranla arka koltukta oturmaya oranla ölüm riskini arttırdığı bulunmuştur (Koşullu lojistik regresyon modeli için $p < 0.05$, sınırlandırılmış lojistik regresyon modeli için $p < 0.01$) (23, 42, 54, 55).

Uygulamada kullanılan otomobilin çarpma hızı grup düzeyinde (otomobil düzeyinde) bağımsız değişken olduğundan, bu bağımsız değişkenin ölüm riski üzerine etkisi sadece sınırlandırılmış lojistik regresyon modelinde araştırılmıştır. Sonuçta otomobilin çarpma hızı ölüm riski üzerine etkili bulunmuştur ($p < 0.01$). O'Neill ve Barry (1995) çalışmasında otomobilin hızı arttıkça bireylerin ölüm riskinin de arttığını belirlemiştir (42).

Özetle, uygulama sonuçlarında da anlaşılacağı üzere emniyet kemeri takmak önden çarpma sonucunda oluşan bir kazada ölüm riskini önemli düzeyde azaltmaktadır. Trafik kazalarında meydana gelen ölüm oranını azaltmak için mutlaka emniyet kemeri takılmalıdır. Bireylerin otomobil içerisindeki konumları da önden çarpma sonucu oluşan bir kazada ölüm riski üzerine önemli bir etkiye sahiptir. Otomobilin ön tarafında oturan bireyler arka tarafta oturan bireylere oranla bu tip kazalarda daha yüksek ölüm riski taşımaktadır. Önden çarpma sonucu oluşan trafik kazalarında ölüm riskini etkileyen bir diğer faktör de otomobilin çarpma hızı olmaktadır. Otomobilin hızı arttıkça araç içerisindeki bireylerin ölüm riskleri de artmaktadır. Bu nedenle trafik kurallarında belirtilen hız limitleri aşılmamalıdır. Ancak yapmış olduğumuz bu çalışmada yaş faktörünün bu tarz kazalarda ölüm riski üzerine her hangi bir etkisinin olmadığı ortaya konulmuştur. Ancak, araç sürücüsünün yaşının ölüm riski üzerine etkisi araştırılmalıdır. Sürücünün yaşı analize alınan verilerde mevcut olmadığından, araç sürücüsünün yaşının ölüm riski üzerine etkisi araştırılmamıştır.

Gomez (2000), hava yastığı kullanımının ölüm riskini önemli derecede azaltacağını belirtmiştir (23). Ancak bu araştırmada bireylerin cinsiyeti ve araçtaki hava yastıklarının ölüm riski üzerine etkileri araştırılmak istenmiş fakat yeterli veri bulunamadığı için önden çarpma sonucu oluşan kazalarda ölüm riski üzerine etkileri belirlenememiştir.

Trafik kazalarında ölüm riski üzerine etki eden faktörler arasında sürücü belgesi süresinin, kaza anında sürücünün arabayı kullanma süresinin, kazanın şehir içi ya da şehirler arası yolda olmasının da önemli etkileri olabileceği düşünülmektedir.

Sağlık ve yaşam üzerine etki eden faktörlerin neden sonuç ilişkilerinin doğru biçimde ortaya konması için sağlıklı ve yeterli verilere gereksinim vardır.

Model seçiminde ise, uygulamada da görüldüğü gibi sınırlandırılmış lojistik regresyon modeli koşullu lojistik regresyon modeline göre tercih edilen model olmalıdır. Çünkü analizlerde ister birey düzeyinde isterse grup düzeyinde bağımsız değişkenler olsun, sınırlandırılmış lojistik regresyon modeli koşullu lojistik regresyon modelinin yerine kullanılabilir.

KAYNAKLAR DİZİNİ

- 1- AGRESTI, A. : An Introduction to Categorical Data Analysis, 103-135, 229-234, John Wiley & Sons, Inc. New York, (1996).
- 2- AKTAŞ, Z., ÖNCÜL, H., URAL, S. : Sayısal Çözümleme, ODTÜ, Ankara, (1991).
- 3- ALBERT, A. : On the Existence of Maximum Likelihood Estimates in Logistic Regression Models, Biometrika, 71: 1-10, (1984).
- 4- ANDERSON, J. A. : Separate Sample Logistic Discrimination, Biometrika, 59: 19-35, (1972).
- 5- BAL, C. : Tedavi Sonrası İzlem Verilerinin Cox Regresyon Aracılığı İle İncelenmesi, (Yayınlanmamış Yüksek Lisans Tezi), Osmangazi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Eskişehir, (1997).
- 6- BANKOLE, A., DARROCH, J. E., SINGH, S. : Determinants of Trends in Condom Use in the United States, Family Planning Perspectives, 31: 264-271, (1999).
- 7- BARR, D., SHERRILL, E. T. : Mean and Variance Truncated Normal Distributions, The American Statistical, 53: 357-361, (1999).
- 8- BHATTACHARYA, P. K., CHERNOFF, H., YANG, S. S. : Nonparametric Estimation of the Slope of a Truncated Regression, The Annals of Statistics, 11: 505-514, (1983).
- 9- BLEIKER, E. M. A., VAN DER PLOEG, H. M. : Personality Factors and Breast Cancer Development: A Prospective Longitudinal Study, Journal of the National Cancer Institute, 88:1478-1483, (1996).

- 10- CHEN, C. H., TSAI, W. Y., CHAO, W. H. : The Product-Moment Correlation Coefficient and Linear Regression for Truncated Data, Journal of the American Statistical Association, 91: 1181-1186, (1996).
- 11-CICIRELLI, V. G., MACLEAN, A. P., COX, L. S. : Hastening Death: A Comparison of Two End-Of-Life Decisions, Death Studies, 24: 401-419, (2000).
- 12- CORCORAN, C., COULL, B., PATEL, A. : Cytel Software Corporation, EGRET for WINDOWS Version 2.0.3, User Manual, 114-154, (1999).
- 13-DAY, N. : Exposure to Power-Frequency Magnetic Fields and the Risk of Childhood Cancer, The Lancet, 354: 1925-1931, (1999).
- 14- DIGGLE, P. J., LIANG, K. Y., ZEGER, S. L. : Analysis Longitudinal Data, 169-189, Oxford University Press, Inc., New York, (1994).
- 15- DORGAN, J. F., LONGCOPE, C. : Serum Sex Hormone Levels are Related to Breast Cancer Risk in Postmenopausal Women, Environmental Health Perspectives Supplements, 105: 583-586, (1997).
- 16- DRAPER, N. R., SMITH, H. : Applied Regression Analysis, John Wiley & Sons, Inc., New York, (1966).
- 17- EDİZ, B. : Lojistik Regresyon-Ayırma Analizi, Ayrımsama Sorunu ve Kalp Hastalarında Lojistik Model Yardımıyla Risk Ölçütlerinin Belirlenmesi, (Yayınlanmamış Doktora Tezi), Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Bursa, (1997).
- 18- EFRON, B., PETROSIAN, V. : Nonparametric Methods for Doubly Truncated Data, Journal of the American Statistical Association , 94: 824-834, (1999).
- 19- ELIASON, S. R. : Maximum Likelihood Estimation: Logic and Practice, Sage Publication, Inc., California, (1993).

- 20- FEINSTEIN, A. R. : Multivariate Analysis: An Introduction, 297-312, 318, 325-330, Yale University Press, New Haven (1996).**
- 21- FITZMAURICE, G. M., LAIRD, N. M., ZAHNER, G. E. P. : Multivariate Logistic Models for Incomplete Binary Responses, Journal of the American Statistical Association, 91: 99-108, (1996).**
- 22- GELFAND, A. E., SMITH, A. F. M., LEE, T. : Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, Journal of the American Statistical Association, 87: 523-532, (1992).**
- 23- GOMEZ, M. S. : Driver Air Bag Effectiveness by Severity of the Crash, American Journal of Public Health, 90: 1575-1581, (2000).**
- 24- GROSS, S. T., LAI, T. L. : Nonparametric Estimation and Regression Analysis with Left-Truncated and Right-Censored Data, Journal of the American Statistical Association, 91: 1166-1180, (1996).**
- 25- GUJARATI, D. N. : Temel Ekonometri, (Çeviri Editörü, ŞENESEN, Ü., ŞENESEN, G. G.), 554-562, Literatür Yayıncılık, İstanbul, (1999).**
- 26- GÜRLER, U. : Bivariate Estimation with Right-Truncated Data, Journal of the American Statistical Association , 91: 1152-1165, (1996).**
- 27- HILL, L. R., HAMMOND, W. G. : Logistic Regression Analysis in Experimental Carcinogenesis, American Statistician, 45: 184-187, (1991).**
- 28- HODOSHIMA, J. : The Effect of Truncation on the Identifiability of Regression Coefficients, Journal of the American Statistical Association, 83: 1055-1056, (1988).**
- 29- KAMO, Y. : Racial and Ethnic Differences in Extended Family Households, Sociological Perspectives, 43: 211-229, (2000).**

- 30- KEY, K. K., DENOON, D. J., BOYLES, S. :** Infection Outbreaks Found in the U.S. and Canada, *Disease Weekly Plus*, 15-17, (1996).
- 31- LANGHOLZ, B., GOLDSTEIN, L. :** Conditional Logistic Analysis of Case-Control Studies with Complex Sampling, *Biostatistics.*, 0: 1-22, (2001).
- 32- LI, G., QIN, J., TIWARI, R. C. :** Semiparametric Likelihood Ratio-Based Inferences for Truncated Data, *Journal of the American Statistical Association*, 92: 236-245, (1997).
- 33- LI, G. :** Nonparametric Likelihood Ratio Estimation of Probabilities for Truncated Data, *Journal of the American Statistical Association*, 90: 997-1003, (1995).
- 34- MAHIEU, L. M., DE DOOY, J. J., LEYS, E. :** Obstetricians' Compliance with CDC Guidelines on Maternal Screening and Intrapartum Prophylaxis for group B Streptococcus, *Journal of Obstetrics and Gynaecology*, 20: 460-464, (2000).
- 35- MANNING W. D., LONGMORE, M. A., GIORDANO, P. C. :** The Relationship Context of Contraceptive Use at First Intercourse, *Family Planning Perspectives*, 32: 104-110, (2000).
- 36- MARSH, G. M., STONE, R. A. :** A Case-Control Study of Lung Cancer Mortality in Four Arizona Smelter Towns, *Archives of Environmental Health*, 53: 15-29, (1998).
- 37- MEHTA, C. R., PATEL, N. R., SENCHAUDHURI, P. :** Efficient Monte Carlo Methods for Conditional Logistic Regression, *Journal of the American Statistical Association*, 95: 99-108, (2000).
- 38- MENARD, S. :** *Applied Logistic Regression Analysis*, 4-5, 12-24, 37-44, 49-52, 80-88, Sage Publications, Inc., Thousand Oaks, (1995).

- 39-** MINITAB Inc. : Minitab User's Guide 2: Data Analysis and Quality Tools, Release 13, Bölüm 2, 29-57, USA, (2000).
- 40-** MONTGOMERY, D. C., PECK, E. A. : Introduction to Linear Regression Analysis, John Wiley & Sons, Inc. New York, (1982).
- 41-** NORUSIS, M. J. : SPSS Regression Models 10.0, 1-14, SPSS Inc., Chicago (1999).
- 42-** O'NEILL, T. J., BARRY, S. C. : Truncated Logistic Regression, Biometrics, 51: 533-541, (1995).
- 43-** ÖZDAMAR, K. : Paket Programlar İle İstatistiksel Veri Analizi-1, 2. Baskı, 475-509, Kaan Kitabevi, Eskişehir, (1999).
- 44-** PAT, J. K., KAPADIA, C. H., OWEN, D. B. : Handbook of Statistical Distribution, 37, 82, 114-115, 220, Marcel Dekker, Inc., New York, (1976).
- 45-** PHILLIPS, C. D., SPRY, K. M., SLOANE, P. D. : Use of Physical Restraints and Psychotropic Medications in Alzheimer Special Care Units in Nursing Homes, American Journal of Public Health, 90: 92-96, (1999).
- 46-** POWELL, J. L. : Symmetrically Trimmed Least Squares Estimation for Tobit Models, Econometrica, 54: 1435-1460, (1986).
- 47-** PRATT, J. W. : Concavity of the Log Likelihood, Journal of the American Statistical Association, 103-106, (1981).
- 48-** QUINN, K. : The Newton Raphson Algorithm for Function Optimization, Washington, (2001).
- 49-** SAS/STAT Software: Changes and Enhancements Through Release 6.12, 875-878, 892-896, 914-918, SAS Institute Inc., North Carolina, (1997).

- 50- SCHNATTER, A. R., ARMSTRONG, T. W. : The Relationship Between Low-Level Benzene and Leukemia in Canadian Petroleum, Environmental Health Perspectives Supplements, 104: 1375-1380, (1996).**
- 51- SPIEGELMAN, D., ROSNER, B., LOGAN, R. : Estimation and Inference for Logistic Regression with Covariate Misclassification and Measurement Error in Main Study/Validation Study Designs, Journal of the American Statistical Association, 95: 51-61, (2000).**
- 52- SPSS Inc. : SPSS Statistical Algorithms 2nd Edition, 138-145, Chicago, (1991).**
- 53- STEWART, M. : Oral Contraceptive Use does not Raise Breast Cancer Risk for Women Older than 40, Canadian Data, Family Planning Perspectives, 25: 142-144, (1993).**
- 54- WEISS, A. A. : A Bivariate Ordered Probit Model with Truncation: Helmet Use and Motorcycle Injuries, Royal Statistical Society, 42: 487-499, (1993).**
- 55- WEISS, A. A. : The Effect of Helmet Use on the Severity of Head Injuries in Motorcycle Accidents, Journal of the American Statistical Association, 87: 48-56, (1992).**
- 56- WILSON, P. D., LANGENBERG, P. : Usual and Shortest Confidence Intervals on Odds Ratios from Logistic Regression, The American Statistical, 53: 332-335, (1999).**

EK SINIRLANDIRILMIŞ LOJİSTİK REGRESYON ANALİZİNİ HESAPLAYAN PROGRAM

Sınırlandırılmış lojistik regresyon analizini hesaplayan ve parametre tahminleri veren program MATLAB 5.3 programlama dili kullanılarak yazılmıştır.

Bu programda kullanılan dosyalar ve semboller aşağıda verilmiştir.

grup.dat: Grup sayısını içeren dat uzantılı dosya.

k.dat: Bağımsız değişken sayısını içeren dat uzantılı dosya.

ni.dat: Gruplardaki birim sayısını içeren dat uzantılı dosya.

yij.dat: Birimlerin durumlarını içeren dat uzantılı dosya.

xij.dat: Bağımsız değişkenleri içeren dat uzantılı dosya.

N: Grup sayısı.

NI(I): Gruplardaki birim sayılarını içeren tek boyutlu dizi (array).

Y(I,J): yij.dat dosyasındaki birimlerin durumlarını içeren verilerin atandığı iki boyutlu dizi (array).

K: Bağımsız değişken sayısı.

XIJ(I,J,L): xij.dat dosyasındaki birimlere ait bağımsız değişkenleri içeren verilerin atandığı üç boyutlu dizi (array).

TAHMIN(I): Parametre tahminlerini içeren tek boyutlu dizi (array).

TERSINFO: Fisher information matrisinin tersini gösteren matris.

SCORE: Parametre tahminlerinin score fonksiyonundaki deęerleri ieren stun vektr.

WALD_ISTATISTIKLERI: Parametre tahminlerine ait Wald istatistiklerini ieren satır vektr.

ODDS_ORANLARI: Parametre tahminlerine ait ODDS oranlarını ieren satır vektr.



BU PROGRAM;
'SINIRLANDIRILMIŞ LOJİSTİK REGRESYON MODELİ İÇİN';
'PARAMETRE TAHMİNLERİNİ, WALD İSTATİSTİKLERİNİ VE ODDS ORANLARINI
HESAPLAMAKTADIR';

```
N=textread('grup.dat','%u');

for I=1:N
    [W0]=textread('ni.dat','%u');
    NI(I)=W0(I);
end

[W1]=textread('yij.dat','%u');
TM=0;
M=0;

for I=1:N
    for J=1:NI(I)
        Z=J+M;
        Y(I,J)=W1(Z);
    end
    TM=TM+NI(I);
    M=TM;
end

K=textread('k.dat','%u');
[W2]=textread('xij.dat','%u');
TMM=0;
MM=0;

for I=1:N
    for J=1:NI(I)
        for L=1:K+1
            ZZ=L+MM;
            XIJ(I,J,L)=W2(ZZ);
        end
        TMM=TMM+(K+1);
        MM=TMM;
    end
end

for I=1:K+1
    TAHMIN(I)=0;
end
```

```

for iterasyon=1:30

for I=1:K+1
    B(I)=TAHMN(I);
end

for I=1:N
    T1=0;
    for J=1:NI(I)
        T1=0;
        for L=1:K+1
            T1=T1+(XIJ(I,J,L)*B(L));
        end
        BX(I,J)=T1;
    end
end

for I=1:N
    for J=1:NI(I)
        P(I,J)=exp(BX(I,J))/(1+exp(BX(I,J)));
        Q(I,J)=1-P(I,J);
    end
end

for I=1:N
    C=1;
    for J=1:NI(I)
        C=C*Q(I,J);
    end
    PI(I)=1-C;
end

for I=1:N
    for J=1:NI(I)
        for L=1:K+1
            A(I,J,L)=P(I,J)*XIJ(I,J,L);
        end
    end
end

T2=0;
for I=1:N
    for L=1:K+1
        for J=1:NI(I)
            T2=T2+A(I,J,L);
        end
        UI(I,L)=(PI(I)^-1*T2);
        T2=0;
    end
end

```

```

end
end
for I=1:N
  for J=1:NI(I)
    for L=1:K+1
      BB(I,J,L)=Y(I,J)*XU(I,J,L);
    end
  end
end
end

T3=0;
for I=1:N
  for L=1:K+1
    for J=1:NI(I)
      T3=T3+BB(I,J,L);
    end
    UBT(I,L)=T3-UI(I,L);
    T3=0;
  end
end

T4=0;
for L=1:K+1
  T4=0;
  for I=1:N
    T4=T4+UBT(I,L);
  end
  UB(L)=T4;
end

for I=1:N
  for J=1:NI(I)
    for A=1:K+1
      for L=1:K+1
        S(I,J,A,L)=XU(I,J,A)*XU(I,J,L);
      end
    end
  end
end

for I=1:N
  for H=1:K+1
    for L=1:K+1
      F(I,H,L)=UI(I,H)*UI(I,L);
    end
  end
end
end

```

```

for I=1:N
  for J=1:NI(I)
    for A=1:K+1
      for L=1:K+1
        G(I,J,A,L)=P(I,J)*Q(I,J)*S(I,J,A,L);
      end
    end
  end
end

TOP=0;
for I=1:N
  for A=1:K+1
    for L=1:K+1
      TOP=0;
      for J=1:NI(I)
        TOP=TOP+G(I,J,A,L);
      end
      TOPLAM(I,A,L)=TOP;
    end
  end
end

for I=1:N
  for A=1:K+1
    for L=1:K+1
      CARPIM(I,A,L)=(PI(I)^-1)*TOPLAM(I,A,L);
    end
  end
end

for I=1:N
  for H=1:K+1
    for L=1:K+1
      CARPIM2(I,H,L)=(1-PI(I))*F(I,H,L);
    end
  end
end

for I=1:N
  for A=1:K+1
    for L=1:K+1
      FARK(I,A,L)=CARPIM(I,A,L)-CARPIM2(I,A,L);
    end
  end
end

```

```

SONTOPLAM=0;
for A=1:K+1
  for L=1:K+1
    SONTOPLAM=0;
    for I=1:N
      SONTOPLAM=SONTOPLAM+FARK(I,A,L);
    end
    IB(A,L)=SONTOPLAM;
  end
end

TERSINFO=inv(IB);
TERSINFO
SCORE=UB'
TAHMIN=B'+(TERSINFO*SCORE)

for I=1:K+1
  WALD_ISTATISTIKLERI(I)=TAHMIN(I)^2/TERSINFO(I,I);
end

WALD_ISTATISTIKLERI

ODDS_ORANLARI(1)=0;
for I=2:K+1
  ODDS_ORANLARI(I)=exp(TAHMIN(I));
end

ODDS_ORANLARI

for A=1:K+1
  for L=1:K+1
    IB(A,L)=0;
  end
end

end

```

ÖZGEÇMİŞ

- 24.07.1975 Eskişehir’de doğdum.
- 1981-1992 Eskişehir’de ilkokul, ortaokul ve lise öğrenimi.
- 1992-1997 Orta Doğu Teknik Üniversitesi, Fen- Edebiyat Fakültesi, İstatistik lisans öğrenimi.
- 1997 (Eylül) Osmangazi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı’nda yüksek lisans eğitimine başlama.
- 1998 (Temmuz) Osmangazi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı’nda Araştırma Görevlisi kadrosuyla göreve başlama.

TC. YÜZÜNCÜYIL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ