T.C.

KARADENİZ TEKNİK ÜNİVERSİTESİ

FEN FAKÜLTESİ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ BÖLÜMÜ

ÇOKLU LOJİSTİK REGRESYON ANALİZİ VE BİR UYGULAMA

BİTİRME TEZİ

HAZIRLAYAN

Doğancan DUMAN-330658

&

Tolga AÇGÜL-314485

Tezin Bölüme Verildiği Tarih : 19.06.2020

Tezin Savunma Tarihi : 10.07.2020

Tez Danışmanı	:	Prof. Dr. Türkan ERBAY DALKILIÇ
Jüri Üyesi	:	Prof. Dr. Zafer KÜÇÜK
Jüri Üyesi	:	Dr. Öğr. Üyesi Uğur ŞEVİK

Trabzon 2020

ÖNSÖZ

“ÇOKLU LOJİSTİK REGRESYON ANALİZİ ve BİR UYGULAMA” isimli bu tez Karadeniz Teknik Üniversitesi Fen Fakültesi İstatistik ve Bilgisayar Bilimleri Bölümü, Lisans Programı’nda hazırlanmıştır. Bu tezin hazırlanması aşamasında yardımlarını esirgemeyen, bize çalışmamızın her aşamasında deneyim ve bilgileri ile yol gösteren danışman hocamız Prof. Dr. Türkan ERBAY DALKILIÇ’a sonsuz teşekkürlerimizi sunarız.

Son olarak, tüm hayatımız boyunca maddi ve manevi her zaman bizi destekleyen, her adımımızda arkamızda duran ailelerimize sonsuz teşekkürlerimizi sunarız.

Bu tezin, bundan sonraki çalışmalara katkı sağlamasını temenni ederiz.

Doğancan DUMAN & Tolga AÇGÜL

Trabzon 2020

İÇİNDEKİLER

Lisans Tezi

ÖZET

ÇOKLU LOJİSTİK REGRESYON ANALİZİ VE BİR UYGULAMA

Doğancan DUMAN & Tolga AÇGÜL

Karadeniz Teknik Üniversitesi
Fen Fakültesi
İstatistik ve Bilgisayar Bilimleri Bölümü
Danışman: Prof. Dr. Türkan ERBAY DALKILIÇ
2020

Regresyon yöntemi bir bağımlı değişken ile bir ya da daha fazla bağımsız değişken arasındaki ilişkiyi çeşitli amaçlar (kestirim, bağımlı değişkeni etkileyen önemli değişkenleri belirleme vb.) çerçevesinde incelemek için kullanılır. Kategorik değişkenin üç ya da üçten fazla olduğu durumlarda çoklu lojistik (multinominal) regresyon analizi kullanılır. Bu analiz iki kategorili (binary) lojistik regresyon analizinin genişletilmiş halidir.

Bu çalışmada çoklu lojistik regresyon analizi ağırlıklı olmak üzere lojistik regresyon analizi ayrıntılı olarak açıklanmıştır. Ayrıca kadın–erkek ilişkisi üzerine bir uygulama yapılmıştır.

Anahtar Kelimeler: Lojistik regresyon analizi, çok kategorili (multinominal) lojistik regresyon analizi.

ŞEKİL ÇİZELGESİ

Sayfa No

SEMBOLLER DİZİNİ

OLS	:	Ordinary Least Squares
MLE	:	Maximum Likelihood
EKK	:	En Küçük Kareler
ÇLR	:	Çoklu Lojistik Regresyon

1. GENEL BİLGİLER

1.1. GİRİŞ

İstatistiksel yöntemlere sağlıkta, meteorolojide, eğitimde, göç olaylarında ve birçok alanda sıkça başvurulmaktadır. Bu yöntemlerden biri de regresyon analizidir. Regresyon yöntemi bir bağımlı değişken ile bir ya da daha fazla bağımsız değişken arasındaki ilişkiyi çeşitli amaçlar (kestirim, bağımlı değişkeni etkileyen önemli değişkenleri belirleme vb.) çerçevesinde incelemek için kullanılır. Amaç verilen veriyi uygun bir modele sokabilmektir. Yani en az değişken ile en iyi uyuma sahip olacak biçimde tanımlayabilen, kabul edilebilir model kurmaktır.

Basit ve çoklu doğrusal regresyonda bağımlı değişken sayısal veri tipindedir. Buna karşılık, çalışmalarda bağımlı değişkenin kategorik/niteliksel veri tipinde olması durumu ile de sıklıkla karşılaşılır. Bu durumda, doğrusal regresyonda parametre kestirimlerini hesaplamak için kullanılan EKK (En Küçük Kareler) yöntemi, bağımlı değişkenin kategorik/niteliksel veri tipinde olması durumu için varsayımlar sağlanmadığından uygun olmamaktadır. Bu nedenle, bağımlı değişken iki ya da ikiden çok kategorili niteliksel veri tipinde olduğunda lojistik regresyon yöntemi ile çözümleme gerçekleştirilir (Alpar & Karabulut, 2017).

Lineer regresyonda bağımsız değişkenlerin parametreleri Ordinary Least Squares (OLS) yöntemiyle tahmin edilirken lojistik regresyonda parametreler Maximum Likelihood (MLE) yöntemiyle hesaplanır. MLE yönteminin amacı sorunsuz parametre havuzundan veri setinin görülme olasılığını maksimize eden en iyi parametreleri seçmek (Şimşek, 2018).

Niteliksel bağımlı değişkenin kategori sayısına ve kategorilerin sırasız (nominal) ya da sıralanabilir (ordinal) olmasına göre farklı lojistik regresyon yöntemleri vardır. Bağımlı değişkenin iki kategorili/durumlu (binary) niteliksel değişken olması durumunda iki kategorili (binary) lojistik regresyon yöntemi kullanılmaktadır. Bağımlı değişken ikiden çok kategorili sırasız (multinominal) niteliksel değişken türünde olduğunda ise çok kategorili (multinomial) lojistik regresyon yöntemi kullanılırken, bağımlı değişken ikiden çok kategorili sıralanabilir (ordinal) niteliksel değişken türünde olduğunda sıralı (ordinal) lojistik regresyon yöntemi kullanılmalıdır (Alpar & Karabulut, 2017).

1.2. BAĞIMSIZ DEĞİŞKENLER

Bir regresyon modelinde birden fazla bağımsız değişken bulunabilir. Bu değişkenler açıklayıcı değişken olup, bağımlı değişkenin değerini tahmin etmek için kullanılır. Regresyon modelinde bu değişkenler ‘x’ parametresi ile temsil edilirler. Çoklu lojistik regresyon modelinin tıp alanına ilişkin uygulamalarında bağımsız değişkenler, bir hastalığın ortaya çıkıp çıkmamasını belirleyen değişkenler ya da risk değişkenleri/faktörleri olarak adlandırılır. Bağımsız değişkeni daha iyi anlamak için şu küçük örneği verebiliriz. Akciğer kanserine yakalanıp yakalanmamayı belirleyen bir değişken, bireylerin sigara içip içmemesidir. Dolayısıyla sigara içmek ya da içmemek bağımlı değişken olan kanseri belirleyen bir risk/bağımsız değişkenidir.

Doğrusal regresyonda olduğu gibi lojistik regresyonda da bağımsız değişkenler niceliksel ya da niteliksel veri türünde olabilmektedir.

1.3. LOJİSTİK REGRESYON İLE DİĞER DEĞİŞKENLİ YÖNTEMLER ARASINDAKİ İLİŞKİ

Lojistik regresyon analizi, bir regresyon yöntemi olmakla birlikte çok değişkenli normallik ve varyansların homojenliği gibi varsayımları olan ayırma (diskriminant) analizine bir alternatif yöntem olarak “ayırma” amaçlı da kullanılmaktadır.

Çoklu doğrusal regresyonda, elde edilen çoklu doğrusal regresyon denklemindeki bağımsız değişkenlere ilişkin regresyon katsayıları/ağırlıkları yardımıyla bağımlı değişkenin gerçek değeri tahmin edilirken, lojistik regresyonda bağımlı değişken kategorilerinden birine atanma olasılığı elde edilir. Dolayısıyla bağımlı değişkene ilişkin kestirimler 0-1 arasında değişir (Alpar & Karabulut, 2017).

1.4. BAĞIMLI DEĞİŞKENİN KODLANMASI

Bağımlı değişkenler, diğer bağımsız değişkenler tarafından etkilenen değişkendir. İkili lojistik regresyon analizinde bağımlı değişkende riskli durumların: 1, risksiz durumların: 0 ile kodlanması yararlı olur. Örneğin bir öğrencinin sınavı geçememesi 1, geçmesi 0 ile kodlanabilir.

Başka bir örnekle, kadınların aile içi şiddet görmesini hangi faktörlerin etkilediği incelenmek istensin. Bağımsız değişkenler için evlilik süresi (yıl), ekonomik sorunlar (yok: 0, az: 1, çok: 2), kötü alışkanlıklar (var: 1, yok: 0), eşin eğitim düzeyi, kadının yaşı gibi değişkenlerin hangisinin, kadının aile içi şiddet görmesini etkileyen önemli faktörler iki kategorili lojistik regresyon ile incelenebilir. Bağımlı değişken için şiddet görmesi 1, görmemesi 0 olarak kodlanır (Alpar & Karabulut, 2017).

2. İKİ KATEGORİLİ (BINARY) LOJİSTİK REGRESYON

Bağımlı değişkenin iki kategorili olduğu lojistik regresyon yöntemidir. Yukarıda da belirtildiği gibi kodlama yapılırken, sıklıkla riskin olmadığı durum için 0 ve riskin olduğu durum için 1 kodu kullanılır. Bağımsız değişkenlerin türü ile ilgili herhangi bir kısıtlama yoktur. Bağımsız değişkenler sürekli sayısal, kesikli sayısal, sırasız ya da sıralanabilir niteliksel değişken türlerinde olabilir (Alpar & Karabulut, 2017).

Lojistik regresyonda bağımlı değişkenin sonucunun kestirilmesi ile ilgilenilmez. Asıl ilgilendiğimiz nokta bağımlı değişkenin 1 (riskli durum 1 olarak belirlenir) değerini alması olasılığıdır. Elde edilen sonuç olasılık değeri olduğu için sadece 0 ile 1 aralığında değerler alabilir. Bilindiği gibi, klasik doğrusal regresyon yöntemi uygulandığında $-\infty$ ile $+\infty$ arasında değerler alır. Yani bizim lojistik regresyonda bağımlı değişken için olan 0 ile 1 aralığını aşar (Alpar & Karabulut, 2017).

Lojistik regresyon metodunun kullanım amacı herhangi bir model inşa tekniğinin amacı ile aynıdır. Bu da bir bağımlı değişken ile bir açıklayıcı değişkenler takımı arasındaki ilişkiyi tanımlayan en iyi matematiksel modeli bulmaktır. En yaygın bilinen model bağımlı değişkenin sürekli olduğu lineer regresyondur. Lojistik regresyon modelini lineer regresyon modelinden ayıran şey bağımlı değişkenin iki sonuçlu (veya kesikli) olmasıdır. Lojistik ve lineer regresyon arasındaki bu fark hem parametrik model seçimini hem de varsayımları etkiler (Bindak, 2017-2018).

İki değerli (binary) y rastgele cevap değişkeni π olasılıklı Bernoulli dağılımına sahip ve p tane açıklayıcı değişkenin ikinci gözlem vektörü $X = (x_{1i}, x_{2i}, \dots, x_{pi})$ olsun. X verilmişken Y'nin koşullu ortalaması olan $E(Y|X) = \pi(x_1)$ aşağıdaki denklem ile tanımlanır (Bindak, 2017-2018).

$$\pi(x_i) = p(Y = 1|X = x) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}}} = \frac{1}{1 + e^{[-(\beta]_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi})}} \quad (1)$$

Bu denkleme Lojistik regresyon fonksiyonu adı verilir. Burada β_i'ler bilinmeyen parametrelerdir, ayrıca $E(y_i) = \pi(x_i)$ olduğundan x_i'nin tüm değerleri için koşullu ortalama [0, 1] aralığında kalır. Lojistik regresyon fonksiyonuna Lojit adı verilen dönüşüm uygulanırsa parametrelere göre doğrusal olan yeni bir değişken elde edilir (Bindak, 2017-2018):

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_{1i} x_{1i} + \cdots + \beta_p x_{pi} \quad (2)$$

Bu durumda yeni değişken $g(x_i)$, Lojit modeldeki parametreler ile doğrusaldır, üstelik sürekli olup $(-\infty, +\infty)$ aralığında değerler alır. π ile $g(x_i)$ aynı yönde doğrusal olarak artar ve eğer $\pi(x) < 0.5$ ise $g(x)$ negatif, $\pi(x) > 0.5$ ise $g(x)$ pozitif değer alır. Negatif ve pozitif değerlerin anlamı ise farklı kategorileri temsil etmesidir. Yani pozitif olması bağımlı değişkenin bir kategorisini, negatif olması ise diğer bir kategorisini temsil etmektedir.

π : x bilindiğinde bağımlı değişkenin 1 değerini alma olasılığı

2.1. LOJİT MODEL

1 numaralı denklem ile verilen lojistik regresyon doğrusallaştırılmalıdır. Bu işlem ise lojit adı verilen bir dönüşüm ile uygulanır. Temel mantığı ise bir olayın odds değerinin doğal logaritmasının alınmasıdır. Bilindiği gibi bir olayın odds'u $\frac{\pi}{1-\pi}$ ile gösterilir. Bu oran 0 ile ∞ arasında bir değer alır. Lojistik regresyon modelini bağımlı değişkenin odds'u cinsinden gösterimi $\frac{\pi(x_i)}{1-\pi(x_i)} = e^{\beta_0 + \beta_{1i} x_{1i}}$ şeklindedir. Bu odds gösteriminin doğal logaritmasının alınması ise bize aşağıdaki gibi lojit dönüşüm sonucunu verecektir.

$$lojit(\pi(x)) = g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = ln e^{\beta_0 + \beta_{1x}} = \beta_0 + \beta_{1x}$$

(3)

Bu sayede lojit dönüşüm β parametrelerinin doğrusal bir fonksiyonnu oluşturur. Yani lojit ile dönüşüm sergileyen bağımsız değişkenler doğrusaldır diyebiliriz. Lojit model doğrusal regresyon modeline benzer. Doğrusal modelde β, bağımsız değişkenin (x) birimlik değişiminin bağımlı değişkene ne kadarlık bir değişimine yol yol açtığını gösterirken lojistik modelde bağımsızın birimlik değişimi lojistikte ne kadar değişime neden olduğunu gösterir. Lojistik regresyonda artıklar 0 ortalama ve $\pi(1-\pi)$ varyans ile binom dağılım sergiler.

2.2. REGRESYON KATSAYILARININ KESTİRİMİ

Lojistik regresyon modelinde katsayıların kestirimi (tahmini) için genellikle lineer regresyonda olduğu gibi maksimum olabilirlik kestirimi yöntemi kullanılır. $(\mathrm{x_i}, \mathrm{y_i})$ gibi n tane bağımsız gözlem eşinin olduğu varsayıldığında, $\mathrm{y_i}$ iki düzeyli sonuç değişkenini, $\mathrm{x_i}$’de i denek için bağımsız değişkenin değerini gösteriyorsa ve sonuç değişkeni için 0 ve 1 kodlarının belirli bir karakteristiğinin yokluğunu ya da varlığını belirlediği kabul edildiğinde lojistik regresyon modelini uydurabilmek için bilinmeyen β_0 ve β_1 parametrelerini kestirmemiz gerekir (Atakurt, 1999).

Lojistik regresyon katsayıları kestiriminde, yinelemeli olmayan ağırlıklandırılmış en küçük kareler, diskriminant fonksiyon analizi yöntemi ve en çok olabilirlik yöntemi (maximum likelihood method) kullanılabilir. Fakat daha çok kullanılan en çok olabilirlik yöntemidir. Bu fonksiyon, gözlenen verinin olasılığını bilinmeyen parametrelerin bir fonksiyonu olarak belirtir. Bu olabilirlik (maximum likelihood) yöntemde bir olayın olması olasılığı maximum yapılmaya çalışılarak bilinmeyen parametrelerin en çok olabilirliği kestirilmeye çalışılır. Örneklem genişliği (n) sonsuza yaklaştıkça, en çok olabilirlik kestiricisi normal dağılım göstererek tutarlılık sergiler. Örneklem küçük olduğu halde iyi sonuçlar da verilebilmektedir. Ama bunun için değişken sayısının en az 10 katı büyüklüğünde, gözlem verisi olması önerilir.

2.3. PARAMETRE TAHMİN YÖNTEMLERİ

Lineer regresyonda bilinmeyen parametreleri tahmin etmek için en çok kullanılan yöntem en küçük kareler yöntemidir. Bu yöntemde modelden tahmin edilen ile gözlenen değerler arasındaki sapmaların (hataların) kareleri toplamını minimize eden β_0 ve β_1 değerleri seçilir. Lineer (doğrusal) regresyon için bilinen varsayımlar altında en küçük kareler yöntemi istenen istatistiki özelliklere sahip tahmin edicileri verir. Ancak, lojistik regresyon analizinde en küçük kareler yöntemi uygulandığı zaman tahmin ediciler aynı özelliklere sahip olmazlar (Hosmer & Lemeshow, 2000). Lojistik modelin parametrelerinin tahmini için kullanılan yöntemler (Şahin, 1999);

- En Çok Olabilirlik Yöntemi
- Yeniden Ağırlıklandırılmış İteratif EKK Yöntemi
- Minimum Lojit Ki-Kare Yöntemi

2.4. KATSAYILARININ ÖNEMLİLİĞİNİN TEST EDİLMESİ

Katsayı kestirimleri elde edildikten sonra modeldeki değişken/katsayı önemliliği test edilir. Modeldeki bağımsız ile bağımlı(yanıt) değişken arasındaki ilişkinin önemliliği incelenir. Bu durum bağımsız değişkenin modelde olduğu ve olmadığı durumun incelenmesi ile önemli olup olmadığı varsayımına gidilir.

Lojistik regresyonda modelin tümel anlamlılığı Olabilirlik Oranı Testi ile yapılırken modeldeki değişkenlerin anlamlılığı Wald ya da Skor (Lagrange multiplier) testlerinden biri ile incelenir. Bu testler aşağıdaki başlıklarda incelenmiştir (Alpar & Karabulut, 2017).

2.4.1. Olabilirlik Oranı Testi

Lojistik regresyon modelindeki bağımsız değişkenin (ya da çoklu lojistik regresyon modelinde bağımsız değişkenlerin veya modelin) anlamlılığı olabilirlik oranı (likelihood ratio-LR) G istatistiği ile incelenir.

$$LR = G = -2\ln\left(\frac{\textit{Değişken modelde olmadığında olabilirlik}}{\textit{Değiken modelde olduğunda olabilirlik}}\right) \tag{4}$$

LR = G = -2(ln(değişken modelde olmadığında L)-ln(değişken modelde olduğunda L))

LR asimtotik olarak ki-kare dağılır. Serbestlik derecesi, iki modelde kestirilen parametre sayısı arasındaki farka eşittir. Bu test olabilirlik oranı testi ya da sapma (deviance) testi olarak adlandırılır. Sapma doğrusal regresyondaki hata kareler toplamına karşılık gelmektedir.

Basit olarak sadece tek bağımsız değişken olması durumu ele alındığında, öncelikle sadece sabit terimin olduğu model oluşturulur. Sonra sabit terimi ile birlikte bağımsız değişkenin olduğu model oluşturulur. Bu modellerden elde edilen iki değer arasındaki fark -2 ile çarpılarak olabilirlik oranı test değeri hesaplanır.

Elde edilen test değerinin küçük olması modele eklenen değişkenlerin lojit'in kestiriminde önemli bir gelişme/katkı sağlamadığını ve değişkenlerin modelde bulunmasına gerek olmadığını gösterir. Bu test işlemi modeli uydurmak için gözlem sayısı (n) yeterince büyük olduğunda geçerlidir.

Olabilirlik oranı testinde modelleri karşılaştırabilmek için tüm modellerin aynı veri setinde karşılaştırılması gerekir. Bu nedenle, veride eksik gözlemler bulunması durumunda, yöntem herhangi. Bir bağımsız değişkende eksik gözlem olması durumunda bu veriyi atarak (listwise delition) verisi tam olan gözlemlerle işlemi gerçekleştirmektedir.

Wald ve Skor testleri değişkenlerin önemliliğini test etmek için kullanılan diğer iki yöntemdir.

2.4.1.1 Wald Testi

Wald testinde de olabilirlik oran testinde olduğu gibi beta katsayılarının en çok olabilirlik kestirimlerinden yararlanılır. Wald testi, eğim parametresi β_1'in en çok olabilirlik kestiriminin ($\hat{\beta}_j$ ya da b_j) standart hatasına ($[\![S(\hat{\beta}]\!]_j)$ ya da $[\![S(b]\!]_j)$) bölünmesi ile elde edilir (Eşitlik 5). Yokluk hipotezi altında Wald test istatistiği standart normal dağılıma uyar.

$$W = \hat{\beta}_j/(S(\hat{\beta}_j)) \tag{5}$$

Yine, β_1'in en çok olabilirlik kestiriminin karesi standart hatasının karesine bölündüğünde Wald istatistiği ki-kare dağılımına uymaktadır (Eşitlik 6)

$$W = \hat{\beta}_j / (\llbracket(S(\hat{\beta}\rrbracket_j)) \sim \chi^2 \tag{6}$$

Wald istatistiği ile ilgili olarak SPSS ve SAS yazılımları ki-kare, STATA yazılımı standart normal dağılım testi sonucu vermektedir.

Büyük örneklemlerde olabilirlik oran testi ve Wald testinin asimptotik olarak benzer sonuçlar verdiği belirtilmektedir. Küçük örneklemlerde hangi testin daha iyi sonuç verdiği konusunda kuramsal bilgiler yetersiz olmakla birlikte, yapılan bazı çalışmalarda wald testi yerine olabilirlik oranı testini kullanılması önerilmiştir.

2.4.1.2 Skor Testi

Skor testinde en çok olabilirlik kestiriminin hesaplanmasına gereksinim yoktur. Bu nedenle, en büyük avantajı hesaplama işlemlerini çok fazla kısaltmasıdır. Genel olarak skor testi matris hesaplamaları gerektiren çok değişkenli bir testtir. Aş. Formül ile belirtilir.

$$ST = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sqrt{y_i (1-\bar{y}) \sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{7}$$

Skor test istatistiği standart normal dağılıma uymaktadır. Bazı istatistik yazılımlarında olmaması bu testin kullanımını sınırlamaktadır.

Skor testinde hesaplama işlemleri çok daha hızlı yapıldığından bazı istatistik yazılımında (SPSS) adımsal yöntemlerde modele alınacak ya da modelden çıkartılacak bağımsız değişkenleri belirlemek için skor testi kullanılmaktadır. SAS yazılımında her üç teste ilişkin sonuçlar edilebilmektedir.

Kat sayılara ilişkin standart hata formülleri ile güven aralığı formülleri burada verilmeyip uygulamalarda tartışılacaktır.

2.5. BAĞIMSIZ DEĞİŞKENLERİN LOJİSTİK REGRESYON MODELİNE ALINMASI

2.5.1. Sayısal (Niceliksel) Değişkenler

Lojistik regresyon modelinde sayısal bağımsız değişkenler modele olduğu gibi alınır. Ek herhangi bir işlem yapılmasına gerek yoktur. Sayısal değişkenin katsayısı bağımsız değişkendeki bir birim değişimin lojit'te ne kadarlık bir değişime neden olduğunu gösterir. Bu yorumun yapılabilmesi için lojit ile sayısal bağımsız değişken arasındaki ilişkinin doğrusal olduğu varsayımının sağlanması gerekir. Bur varsayımı incelemek için çeşitli grafiksel yöntemlerden ya da testlerden yararlanılabilir. Bu yöntemler arasında işlemlerin basit olması nedeni ile Box-Tidwell test yaklaşımı. En çok tercih edilendir.

Box-Tidwell yaklaşımında, sayısal bağımsız değişken x ile bu değişkenin doğal logaritması ln(x) çarpılarak, xln(x) etkileşim terimi modele eklenir. Etkileşim teriminin modelde önemli bulunması lojit ile sayısal bağımsız değişken arasındaki ilişkinin doğrusal olmadığını belirtir. Varsayım sağlanmadığında, bağımsız değişkende dönüşüm yapılarak ilişkinin doğrusallaştırılması ya da sayısal değişkenin sınıflanarak niteliksel değişken olarak modele alınması seçeneklerinden birisi uygulanabilir.

Box-Tidwell testinin doğrusallıktan küçük ufak ayrılışları belirlemede gücünün küçük olduğu belirtilmektedir. Bu nedenle Hosmer ve Lemeshow, lojit ile sayısal bağımsız değişken arasındaki ilişkinin doğrusal olup olmadığını incelemek için sayısal değişkenin sınıflandırılmasını önermişlerdir. Bu yöntemde sayısal değişken çeyreklik değerlerine göre dört sınıfa ya da yüzdelik değerlerine göre her sınıfta gözlem sayısı benzer olacak şekilde beş sınıfa ayrılır. Ve yeni değişken dizayn değişkeni olarak modele eklenir. Sonrasında dizayn değişkeni yardımıyla elde edilen lojit değerleri incelenerek sayısal bağımsız değişken ile lojit arasındaki ilişkinin doğrusal olup olmadığı değerlendirilir. İstenirse dikey eksende lojit değeri ve yatay eksende de sınıflar olacak şekilde grafik çizilerek de değerlendirme yapılır.

2.5.2. Niteliksel Değişkenler

- **İki kategorili niteliksel değişkenler**

İki durumlu/kategorili niteliksel değişkenler lojistik regresyon modelinde olduğu gibi kullanılır. Burada dikkat edilecek nokta yorumlamanın kolay olması açısından, riskin az olduğu düşünülen gruba küçük, riskin fazla olduğu düşünülen gruba büyük kodun verilmesidir. Örneğin bağımsız değişken sigara içmek ise; sigara, içmeyenlere 0, ve içenlere 1 kodu verildiğinde modelden elde edilen katsayı, risk faktörü olarak düşünülen sigara içmenin lojit'te ortaya çıkardığı değişimi göstermektedir. Gerektiğinde aşağıda tanımlanacak olan kontrastlar kullanılarak referans kategori istenilen doğrultuda belirlenebilir. Böylece incelenen değişkendeki kodlar değiştirilmeden referans kategori olarak istenilen kategorinin belirlenmesini sağlar.

- **Sırasız niteliksel (nominal) değişkeler**

Sırasız niteliksel (nominal) değişkenleri modelde kullanabilmek için çalışmanın amacına uygun dizayn (ya da kukla) değişkenleri oluşturması gerekir. Seçilen kontrast tipine bağlı olarak dizayn değişkenlerine (kategori sayısı- 1 tane) istatistiksel yazılımlarının birçoğu kendisi oluşturmaktadır.

3. ÇOKLU LOJİSTİK REGRESYON ANALİZİ

Regresyon analizi yönteminde olduğu gibi lojistik regresyon yönteminde de birden fazla bağımsız değişken, bağımlı değişken durumunu değiştirebilir. Lojistik regresyonda iki veya daha fazla bağımsız değişkenin olduğu durumlar çoklu lojistik regresyon olarak tanımlanır. Çoklu lojistik regresyon ise basit tek değişkenli lojistik regresyonun genişletilmesi ile gerçekleşir. Değişken sayısı p olmak üzere bağımsız değişkenlerin vektörü $x = (x_1, x_2, \dots, x_p)$ bu şekilde verilebilir. Buradan, çoklu lojistik regresyon denklemi aşağıdaki gibi yazılır.

$$\pi(x) = P(Y = 1|x) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_{2+\cdots}+\beta_p x_p)}}{1 + e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}} \quad (8)$$

Lojistik regresyon modeli, yanıt değişkeninin odds'u türünden aşağıdaki gibi belirtilebilir.

$$\frac{\pi(x)}{1-\pi(x)} = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)} \quad (9)$$

Odds'un doğal logaritması alınarak lojit dönüşüm yapılmış olur.

$$lojit\ \pi(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \quad (10)$$

Odds'un doğal logaritması alındığında model doğrusal modele dönüşür ve aşağıdaki gibi gösterilir.

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (11)$$

Çoklu lojistik regresyon modeli kurulduktan sonra modeldeki değişkenlerin önemliliği test edilir. Tek değişkenli durumda olduğu gibi olabilirlik oran testi modeldeki bağımsız değişkenler için p katsayının tümel önemliliğini verir. Diğer bir deyişle, olabilirlik oran testi modeldeki p adet β katsayısının sıfıra eşit olup olmadığını test eder.

Tümel test ile modeldeki katsayıların sıfırdan farklı olduğu sonucu bulunduğunda, Wald testi ile modeldeki her bir değişkenin tek tek önemliliği incelenir. Çoklu lojistik regresyon analizinde de diğer çok değişkenli çözümleme yöntemlerinde olduğu gibi en az değişken ile en iyi uyuma sahip model elde edilmek istenir (Alpar & Karabulut, 2017).

Çoklu lojistik regresyon analizi uygulanmadan önce çalışmada kullanılan bağımsız değişkenlerin tek değişkenli analizler ile incelenerek ilgisiz olanların modele alınmaması önerilmektedir. Bu amaçla, iki yaklaşımdan yararlanılır. Birinci yaklaşım da bağımlı/yanıt/sonuç değişkeni ile ilgili bağımsız değişkenler arasında bir ilişki olup olmadığı ki-kare, iki ortalama arasındaki farkın önemlilik testi, Mann Whitney U testi gibi testlerden yararlanılarak incelenir (Alpar & Karabulut, 2017).

Diğer bir yaklaşım ise her bir bağımsız değişkenle tek değişkenli lojistik regresyon analizi yapılması ve p değerinin incelenmesi gerekir. Her iki yaklaşım sonucunda da (yani tek değişkenli analizler sonucunda) p değeri 0.25'in altında bulunan değişkenlerin çok değişkenli çözümlemelerde dikkate alınması önerilmektedir. Böylece çok değişkenli çözümlemede daha az sayıda değişken ile ilgilenilir. Özellikle gözlem sayısının az olduğu çalışmalarda, bu yaklaşım çoklu bağlantı gibi sorunların ortaya çıkmasını önleyebilir. Bilindiği gibi genellikle çok değişkenli analizlerin iyi sonuç verebilmesi için değişken sayısının an az 10 katı gözlem ile çalışmanın yürütülmesi gerekmektedir (Alpar & Karabulut, 2017).

3.1. ETKİ KARIŞIMI

Hem bağımlı/sonuç değişkeni hem de bağımsız değişken ile ilişkisi bulunan ortak değişkene etki karıştırıcı değişken denir.

Etki karışımını gidermek için olası etki karıştırıcı değişkenin modele eklenmesi gerekir. Örneğin akciğer kanseri üzerinde sigara içmenin etkisinin araştırıldığı bir çalışmada, yaşın ortak değişken (covariate) olarak modele katıldığı varsayılsın. Sigaranın etkisi, sadece sigara bağımsız değişken olarak alınıp hesaplandığında ve sigara ile birlikte yaş grupları (50 yaş altı – 50 yaş ve üzeri) değişkeni bağımsız değişken olarak alınıp hesaplandığında farklılaşıyor; ancak bu fark istatistiksel olarak önemli bulunmuyorsa, yaşın etki karıştırıcı bir değişken olduğu söylenir (Alpar & Karabulut, 2017).

3.2. ETKİLEŞİM

Regresyon modelindeki bir bağımsız değişkenin bağımlı değişken ile ilişkisi, diğer bir bağımsız değişkenin düzeyine göre değişiyorsa etkileşim olduğu söylenir. Yine akciğer kanseri ile sigara içme için örnek verilecek olursa, ailesinde kanser öyküsü olanlarda ve olmayanlarda sigara içenlerin akciğer kanserine yakalanma riski farklıysa ve bu fark istatistiksel olarak önemliyse burada bir etkileşimden söz edilir. Örneğin ailesinde kanser öyküsü olmayıp sigara içenlerde akciğer kanserine yakalanma riski 1,5 kat fazla iken, ailesinde kanser öyküsü olup sigara içenlerde bu risk 3,5 kata çıkıyorsa ve bu değişim istatistiksel olarak anlamlı ise, sigara ile aile öyküsü değişkenleri arasında etkileşim olduğunu gösterir (Alpar & Karabulut, 2017).

Etkileşim terimini içeren lojistik regresyon modelini veriye uydurmak doğrusal regresyon analizindekine oldukça benzerdir. Örneğin x1 ve x2 gibi iki sürekli bağımsız değişken arasındaki etkileşim, x1 ve x2 değişkenlerini içeren modele bu iki değişkenin çarpımı olan x1x2 değişkeni eklenerek incelenir. Bu çarpım olarak eklenen bağımsız değişkenin istatistiksel olarak etkisinin önemli bulunması x1 ve x2 arasında bir etkileşim olduğunu ortaya çıkartır. Sonuçları yorumlayabilmek için etkileşim terimi önemli bulunduğunda, ana etkiler önemsiz bulunsa bile modele alınması gerekir.

3.3. ÇOKLU BAĞLANTI SORUNU

Bağımsız değişkenlerin birbirleriyle ilişkisi olduğu durumda çoklu bağlantı sorunu ortaya çıkar. Eğer bu değişkenler arasında tam ilişki varsa, regresyon modelindeki katsayıları elde edemeyiz. Örnek vermek gerekirse, elimizde iki adet sayısal bağımsız değişkenimiz olsun. Bunların toplamından elde edilecek üçüncü bir değişken modele eklendiği durumda katsayılar hesaplanamaz.

Uygulamada, bağımsız değişkenler arasında mükemmel çoklu bağlantılı durumlarda karşılaşma olasılığı çok düşüktür. Çoklu bağlantı sorunu genellikle değişkenler arasındaki yüksek korelasyon nedeniyle gerçekleşir. Bağımsız değişkenler arasında çoklu bağlantı olduğunda, bu değişkenlerden sadece birinin modele alınması gereklidir.

3.4. Model Seçim Stratejisi

Model oluşturma işleminde temel amaç modelde bulunması gereken faktörlerin ya da değişkenlerin belirlenmesidir. Model seçim stratejisi çalışmanın amacına göre düzenlenir. Bazı uygulamalarda sonuç değişkeninin, hangi bağımsız değişken ile daha çok etkilendiği bulunmak istenebilir. Asıl amaç hangi değişkenlerin bağımlı değişkeni etkilediği araştırılmaktadır.

3.4.1. Değişken Seçim İşlemi

Değişken sayısı az iken etkileşim terimi ve doğrusal olmayan terimleri içeren olası açıklayıcı/bağımsız değişkenlerin sayısı az olacağından, olası tüm modellerin oluşturulması önerilir. Bu modeller arasından uyumu yüksek olan model seçilmelidir. En uygun modelin seçiminde yalnızca istatistiklere bağlı kalınmaması, araştırmacının amacına göre hangi modelin uygun olduğuna karar verebilmelidir.

Açıklayıcı değişken sayısı fazla olduğunda olası tüm modellerin sayısı da fazla olacaktır. Yaklaşık olarak değişken sayısının karesi kadar. Değişken seçimi için böylesi durumlarda adımsal yöntemler kullanılır.

3.4.1.1 İleriye Yönelik Seçim Yöntemi

Bu yöntem için aşağıdaki adımlar izlenir:

- Sadece sabitin olduğu, bağımsız değişkenlerin olmadığı model ile işleme başlanır.
- Modele eklendiğinde log olabilirlik üzerinde en fazla değişime neden olan bağımsız değişken belirlenir ve bu değişken modele alınır.
- Eklenen değişkenin modele katkısı önemsiz bulunana kadar modele bağımsız değişkenler eklenmeye devam edilir.

3.4.1.2 Geriye Dönük Eleme

Bu aşamada bütün değişkenler modele eklenir ve teker teker modelden çıkarılarak, varyans değerleri göz önüne alınır. Varyans değerleri küçük olanlar modelden kalıcı olarak çıkarılır. Varyans değeri yüksek olanlar ise modelde kalmaya devam edilip çıkarılmaz.

3.5. LOJİSTİK REGRESYONDA AÇIKLAYICILIK KATSAYILARI (R^2)

Doğrusal regresyondaki en küçük kareler yöntemi ile R^2 hesabını lojistik regresyon için kullanmak pek doğru bir yöntem değildir. Yani kısacası doğrusal regresyondaki R^2'nin lojistik regresyonda karşılığı yoktur. Lojistik regresyon için modelde uyumun bir göstergesi olarak kullanabilmek için yalancı açıklayıcılık katsayıları (R^2) vardır. Bu katsayı değerleri Doğrusal regresyondan farklı olarak küçük değerler verdiği için uyum açısından dikkat edilmesi gerekmektedir.

Doğrusal regresyonda bağımlı değişkendeki değişimin ne kadarının bağımsız değişken tarafından karşılandığını açıklayabilmek için açıklayıcılık katsayısı (R^2) kullanılır. Doğrusal regresyonda uyum iyiliği ölçüsü kullanılarak R^2'nin hesaplanması aşağıdaki gibidir.

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

(12)

n: gözlem sayısı

y_i: bağımlı değişken

$\bar{y}$: Bağımlı değişken y'nin ortalaması

$\hat{y}_i$: Model yardımıyla kestirilebilen değer.

Lojistik regresyonda uyum iyiliğini değerlendirmek için kullanabileceğimiz çeşitli yalancı R^2 istatistikleri geliştirilmiştir. McFaden R^2, Cox-Snell R^2 ve Nagelkerke R^2 istatistikleri en sık kullanılan istatistiklerdir.

3.5.1. McFaden R^2

McFaden tarafından önerilen (1973), "olabilirlik oranı indeksi" olarak da adlandırılan yöntemde, bağımsız değişkenlerin olmadığı sadece sabitin bulunduğu modelin (M_α) log olabilirliği genel kareler toplamı, bağımsız değişkenlerin olduğu modelin (M_β) log olabilirliği de artık kareler toplamı olarak düşünüldüğü R^2 istatistiği aşağıdaki eşitlik ile verilebilir (Alpar & Karabulut, 2017).

$$R_{McF}^2 = 1 - \frac{ln\hat{L}(M_\beta)}{ln\hat{L}(M_\alpha)} \quad (13)$$

Eğer $M_{\alpha} = M_{\beta}$ olursa (yani tüm eğimler 0 olursa) McFaden R^2 0 değerini alır; ancak hiçbir zaman 1 değerini alamaz

3.5.2. Düzeltilmiş McFaden R^2

Doğrusal regresyonda kullanılan R^2'de olduğu gibi McFaden R^2'de modele değişken eklendiğinde artmaktadır. Bu nedenle, Ben-Akiva ve Lerman (1985) McFaden R^2'nin modeldeki değişken sayısı (p) ile düzeltilmesini önermiştir. Düzeltilmiş McFaden R^2:

$$R^2_{McF} = 1 - \frac{ln\hat{L}(M_{\beta}) - p}{ln\hat{L}(M_{\alpha})} \qquad (14)$$

Böylece sadece modele eklenen her bir değişken $ln\hat{L}(M_{\beta})$'da 1 birimden fazla artışa neden olduğunda düzeltilmiş McFaden R^2 artmaktadır.

McFaden R^2, çoklu doğrusal regresyonda elde edilen R^2'ye göre oldukça küçük değerler alma eğiliminde olduğunda 0.20 ile 0.40 arasındaki bir değerin çok yüksek olduğu söylenir.

3.5.3. Cox-Snell R^2

En çok olabilirlik R^2 olarak da bilinen istatistik Madala tarafından önerilmiştir.

$$R^2_{CS} = 1 - \left\{\frac{ln\hat{L}(M_{\alpha})}{ln\hat{L}(M_{\beta})}\right\}^{\frac{2}{n}} = 1 - e^{(-\frac{G^2}{n})}$$

(15)

Cox-Snell R^2 en küçük 0 değerini alırken en büyük değeri 1 olmamaktadır.

3.5.4. Nagelkerke R^2

Cox-Snell R^2 istatistiğinin alabileceği en büyük değer $1 - \hat{L}(M_{\alpha})^{\frac{2}{n}}$'dir. Nagelkerke'nin önerdiği düzeltme ile Cox-Snell R^2'nin en büyük değeri 1 olabilmektedir:

$$R_{NAG}^2 = \frac{R_{CS}^2}{R_{MAX}^2} = \frac{1-\left\{\frac{\hat{L}(M_\alpha)}{\hat{L}(M_\beta)}\right\}^{\frac{2}{n}}}{1-\hat{L}(M_\alpha)^{\frac{2}{n}}} \tag{16}$$

Nagelkerke R^2 aynı zamanda Cragg ve Uhler R^2 olarak da adlandırılmaktadır.

3.6. MODEL UYUM İYİLİĞİNİN DERECELENDİRİLMESİ

Modelin uyum iyiliği ile lojistik regresyon analizinden elde edilen modelin sonuç değişkenini tanımlamakta kullanılır. Modelin veriye uyumunun iyi olup olmadığı çeşitli yöntemle incelenebilir.

3.6.1. Pearson ki-kare ve Sapma (Deviance) istatistikleri

Gözlenen değerler ve uydurulan model yardımıyla elde edilen beklenen değerler karşılaştırılarak modelin uyumu değerlendirilebilir. Pearson ki-kare testi ve sapma istatistikleri gözlenen ve beklenen değerler arasındaki farkı kullanarak modelin uyumunu değerlendirir. Gözlenen ve kestirilen değerler arasındaki farklar, diğer bir deyişle artıklar kullanılarak bu iki istatistik elde edilir. Pearson artıklarının toplamı Pearson Ki-Kare istatistiğini, sapma artıklarının toplamı sapma istatistiğini verir. Bu iki istatistik sırasıyla aşağıdaki gibi gösterilebilir:

$$\chi^2 = \sum_{i=1}^{J} r_i^2 \tag{17}$$

$$D = \sum_{i=1}^{J} d_i^2 \tag{18}$$

Burada J, x'in ($x = (x_1, x_2, \ldots, x_P)$) gözlenen farklı değerlerini belirtmektedir. Bazı gözlemlerin x değerleri aynı olduğunda J<n olur. Her iki istatistikten elde edilen küçük P değerleri modelin uyumunun kötü olduğunu gösterir. Hem Pearson ki-kare hem de sapma istatistiğinin J-p-1 serbestlik derecesi ile ki-kare dağılımına uyduğu varsayılmaktadır; ancak bazı çalışmalarda $J \approx n$ ise bunun doğru olmadığı gösterilmiştir. Bağımsız

değişkenler içerisinde sürekli sayısal değişkenler olması durumunda x’ler çok farklı değerler alacağından bu sorunla karşılaşılması olasıdır (Alpar & Karabulut, 2017).

3.6.2. Hosmer ve Lemeshow Testi

Hosmer ve Lemeshow’ un uyum iyiliğini incelemek için geliştirdikleri yöntemde lojistik regresyon analizinden kestirilen olasılıklar kullanılır. Kestirilen olasılıklar küçükten büyüğe doğru sıraya dizilir ve bu olasılıklara göre bireyler k alt gruba bölünür. Sıklıkla k değeri 10 olarak kullanılır. Sonrasında, her alt grupta gözlenen ve beklenen değerler hesaplanarak bilinen ki-kare testi uygulanır (Alpar & Karabulut, 2017).

$$\hat{C} = \sum \frac{(G - B)^2}{B} \tag{19}$$

Bu istatistik k-2 serbestlik derecesi ile Ki-kare dağılımına uyar.

Hem klasik yaklaşım hem de Hosmer-Lemeshow’ un yöntemleri için büyük örneklem genişlikleri gerektir. Hosmer-Lemeshow’ un yöntemini uygulamak için en az 400 gözlem olması önerilmektedir. Elde edilen büyük ki-kare değerleri modelin uyumunun iyi olmadığını gösterir.

Hosmer-Lemeshow test istatistiği $\hat{C}$’yi hesaplamak için grup sayısı küçük seçilirse $\hat{C}$ istatistiği gözlenen ve beklenen frekanslar arasındaki farklılığı ayırt etmekte çok duyarlı olmamaktadır. Hosmer ve Lemeshow grup sayısı 6’dan az seçildiğinde $\hat{C}$ istatistiğinin hemen hemen her zaman modelin uyumunun iyi olduğu sonucunu verdiğini belirtmişlerdir.

$\hat{C}$ gibi özet uyum iyiliği testleri kullanmanın avantajı uyumu değerlendirmek için kolay yorumlanabilen tek bir değer elde edilmesidir. Büyük dezavantajı ise gözlem sayısı az olduğunda bu istatistiklerin uyumdaki önemli sapmaları belirleyememesidir. Bu nedenle uyumunun iyi olduğuna karar vermeden önce her bir gözlem için artıkların ve ilgili istatistiklerin incelenmesi önerilir (Alpar & Karabulut, 2017).

3.6.3. Artıkların İncelenmesi

Regresyon modelinin uyumunun incelenmesinde artıkların büyük bir rolü vardır. Doğrusal regresyonda artıklar e simgesi ile gösterilir ve gözlenen değer ile kestirilen değer arasındaki farkı ($e_i = y_i - \hat{y}_i$) belirtir. Doğrusal regresyonda, artıkların dağılımının 0 ortalama ve sabit bir varyansla normal dağılıma uyması ve kestirim değerinden ($\hat{y}_i$) bağımsız olması beklenir. Buna karşılık, lojistik regresyonda artıkların büyüklüğü ve varyansı kestirilen olasılık $\hat{\pi}_i$'nin değerine bağlıdır. Diğer bir deyişle artıklar normal dağılım göstermez ve değişen varyanslılık sorunu vardır. Bu nedenle, lojistik regresyonda artıkların incelenmesi daha karmaşıktır.

Lojistik regresyonda model yeterliliğini incelemede Pearson, standartlaştırılmış pearson artıkları ve sapma (deviance) artıkları sıklıkla kullanılmaktadır. Sapma artıklarının normal dağılıma yakın olması ve $\hat{pi}$'nin değeri 0 ya da 1'e yakın olduğu durumda Pearson artıkların sağlam olmaması/kararsız olması (unstable) nedeniyle alanyazında sapma artıklarının kullanılması yeğlenmektedir. Bunların yanı sıra, lojit artıklar ve student türü artıklar da kullanılmaktadır (Alpar & Karabulut, 2017).

3.6.3.1 Pearson Artıkları (r_i)

Değişen varyanslılık sorunu nedeniyle ham artık değeri $y_i - \hat{\pi}_i$ kendi standart sapmasına bölünerek Pearson artıkları elde edilir:

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}} \qquad (20)$$

r'nin değerinin büyük olması o gözlem için modelin uyumunun iyi olmadığı anlamına gelir. Bu değerin genellikle -2 ile +2 arasında olması beklenir. Pearson değeri bu aralıkta olmayan gözlemler aşırı gözlem olarak nitelendirilir.

3.6.3.2 Standartlaştırılmış Pearson Artıkları (r_{std_i})

Yapılan çalışmalarda artıkların varyansı $\hat{\pi}_i(1 - \hat{\pi}_i)$'ye eşit olmadığında varyansının 1 olmadığı gösterildiğinden standartlaştırılmış Pearson artıklarının kullanılması önerilmiştir.

$$r_{std_i} = \frac{r_i}{\sqrt{1 - h_{ii}}} \tag{21}$$

Burada h_{ii} şapka (hat) matrisinin köşegen elemanlarıdır ve $h_{ii} = \hat{\pi}_i(1 - \hat{\pi}_i)x_i Var(\beta) x_i{}'$ eşitliği ile elde edilir. Aynı sonuca, doğrusal regresyon için verilen formüllerle de ulaşılabilir. Bu yaklaşım sonucunda, standartlaştırılmış Pearson artıklarının varyansı sabitlenmiş olur.

3.6.3.3 Lojit Artıklar (l_i)

Lojit artıklar Pearson artıklarına benzer şekilde hesaplanır. Gözlenen değer y_i ile kestirilen olasılık $\hat{\pi}_i$ arasında fark kendi varyansına bölünerek lojit artıklar elde edilir.

$$l_i = \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)} \tag{22}$$

3.6.3.4 Sapma (deviance) Artıklar (d_i)

Lojistik regresyonda sapma artıklar, her gözlemin sapma istatistiği D'ye olan katkısını belirtir. Burada bahsedilen D, modelin tümel uyum eksikliğini gösteren istatistiktir.

$$d_i = sign(y_i - \hat{\pi}_i)\sqrt{-2[(-y_i \ln(\hat{\pi}_i)) - (1 - y_i)\ln(1 - \hat{\pi}_i)]} \tag{23}$$

Burada, $sign(y_i - \hat{\pi}_i)$ fonksiyonu aşağıdaki değerleri alır;

$$-1 \; eğer \; (y_i - \hat{\pi}_i) < 0$$

$$0 \; eğer \; (y_i - \hat{\pi}_i) = 0$$

$$+1 \; eğer \; (y_i - \hat{\pi}_i) > 0$$

Sapma artıklarının ortalaması kestirilen olasılıklara bağlı olduğundan bu artıkların yorumlanmasında sorun ortaya çıkmaktadır.

3.6.3.5 Standartlaştırılmış Sapma Artıkları (d_{std_i})

Standartlaştırılmış sapma artıkları da şapka matrisi yardımıyla hesaplanır. Standartlaştırılmış sapma artıkları d_{std_i} aşağıdaki formül ile verilir.

$$d_{std_i} = \frac{d_i}{\sqrt{1 - h_{ii}}} \tag{24}$$

3.6.3.6 Student Türü Artıklar

Student türü artıklar, her bir gözlem dışarıda tutularak yapılan regresyon analizi sonuçlarına dayalı olarak hesaplanmaktadır. Hem pearson hem de sapma artıkları için student türü artıklar tanımlanmıştır. Sapma artıklarına dayalı olarak hesaplanan student artıklar asimptotik olarak normal dağılmaktadır. Genellikle -2 ile +2 arasında olması istenir (Alpar & Karabulut, 2017).

3.7. KATSAYILARIN YORUMLANMASI

Bilindiği gibi doğrusal regresyonda eğim katsayısı β_1, bağımlı değişkenin $x + 1$'e karşılık gelen değeri ile x'e karşılık gelen değeri arasındaki farka eşittir. Diğer bir yaklaşımla $y(x) = \beta_0 + \beta_1 x$ olduğunda $\beta_1 = y(x + 1) - y(x)$'e eşittir.

Bezer şekilde lojistik regresyonda eğim katsayısı, bağımsız değişkendeki 1 birim değişime karşılık lojit'te oluşan değişimi gösterir ($\beta_1 = g(x + 1) - g(x)$).

3.7.1. Modelde Niteliksel Bir Bağımsız Değişken Olduğunda

Bağımsız değişken x, 0 ya da 1 değerini alan iki kategorili niteliksel değişken olduğunda lojit 'teki değişim $g(1) - g(0) = [\beta_0 + \beta_1] - [\beta_0] = \beta_1$ olarak bulunur. Dolayısıyla lojit'teki değişim, lojistik regresyon katsayısı β_1 kadar olduğu görülmektedir. Sonuçlar yorumlanırken okuyucular tarafından daha kolay yorumlanabilsin diye genellikle lojit yerine odds oranı kullanılmaktadır. Odds oranı β_1 katsayısının e üssü (exponansiyeli) alınarak (e^{β_1}) hesaplanır.

Örneğin sigara içmenin erken doğum üzerine etkisini incelemek için uygulanan lojistik regresyon sonucunda, sigara değişkeninin katsayısının $\beta_1 = 0.588$ olarak elde edildiğini varsayalım. Bu sonuç, "sigara içen kadınlarda içmeyen kadınlara göre erken doğum riski 1,8 kat ($e^{0,578} = 1,8$) daha fazladır" biçiminde yorumlanır.

3.7.2. Modelde Sürekli Bir Bağımsız Değişken Olduğunda

Bağımsız değişken x sürekli sayısal değişken olduğunda, lojir ile sürekli değişken arasındaki ilişkinin doğrusal olduğu varsayımı altında:

$$g(x+1) - g(x) = [\beta_0 + \beta_1(x+1)] - [\beta_0 + \beta_1 x] = \beta_1 \qquad (25)$$

Eğim katsayısı β_1, bağımsız değişken x'deki 1 birim değişime karşılık lojit'te oluşan değişimi gösterir

Sürekli bağımsız değişkenler için kestirilen katsayıların yorumlanması, bu değişkenlerin modele nasıl alındığına ve değişkenlerin birimine bağlıdır. Çoğu zaman bağımsız değişkendeki 1 birimlik artış biyolojik olarak önemli olmamaktadır. Örneğin yaştaki 1 birim artış önemli sayılmayacak kadar küçük bir değişime neden olabilir. Bunun yerine, yaştaki 10 birimlik artışın neden olduğu değişimi incelemek daha yararlı olabilir.

Sayısal bağımsız değişken 0 ile 1 arasında değer alan oran olarak modele alındığında tam tersi bir durum söz konusudur. Bu defa, bağımsız değişkendeki 1 birimlik artış çok büyük olacaktır. Bağımsız değişkendeki 0,01 birimlik ya da 0,10 birimlik artışa göre yorum yapılması daha gerçekçi olacaktır.

Bağımsız değişkendeki x'deki c birimlik artış için lojitte oluşan fark:

$$g(x+1) - g(x) = c\beta_1 \qquad (26)$$

şeklindedir. Lojitteki bu farkın e üssü alınarak ilgili odds oranı $e^{c\beta_1}$ elde edilir. Odds oranının güven aralığını hesaplamak için gereksinim duyulan standart hata, $\widehat{\beta_1}$ 'in standart hatası c ile çarpılarak elde edilir. Güven aralığı aşağıdaki gibidir.

$$e^{\left(c\widehat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} cS(\widehat{\beta}_1)\right)} \qquad (27)$$

Görüldüğü gibi odds oranı için nokta kestirimi ve güven aralıkları seçilen c değerine bağlı olmaktadır.

3.7.3. Etkileşim Olduğunda Katsayıların Yorumlanması

Modelde etkileşim terimi önemli bulunduğunda katsayıların yorumlanması biraz karmaşıktır. Direkt olarak etkileşim terimine ilişkin katsayının e üssü alınarak yorumlanması hatalıdır. Etkileşim teriminin katsayının üssü alındığında elde edilen sonuç bir odds oranı değildir. Bu değerin yanlışlıkla odds oranı olarak yorumlanmasının ana nedeni istatistik yazılımlarının katsayıların odds oranı olarak vermesidir.

Yine benzer şekilde, etkileşimi dikkate almadan bağımsız değişkenler için odds oranını hesaplamak da hatalıdır. Modele etkileşim terimi önemli bulunduğunda izlenecek en iyi yol ilgilenilen odds oranlarının elle hesaplanmasıdır.

3.7.4. Güven Aralıklarının Hesaplanması

Elde edilen odds oranları için uygun güven aralıklarının hesaplanması biraz daha karmaşıktır. Hosmer-Lemeshow birden fazla kestirilen katsayı kullanılarak hesaplanan odds oranı için parametrelerin kestirilen kovaryansı yardımıyla güven aralıklarının hesaplanmasını göstermiştir. Risk faktörü F, ortak değişken X ve etkileşim terimi $F.X$'i içeren bir model olduğu düşünüldüğünde, $F = f$ ve $X = x$ değerlerinde model için lojit aşağıdaki gibidir.

$$f(f, x) = \beta_0 + \beta_1 f + \beta_2 x + \beta_3 (f.x) \tag{28}$$

X değişkeni $X = x$'de sabit tutulduğunda $F = f_1$'in $F = f_0$'a karşı logodds değeri aşağıdaki formül ile elde edilir.

$$\ln[\Psi(\mathrm{F} = f_0, \mathrm{F} = f_1, \mathrm{X} = \mathrm{x})] = g(f_1, x) - g(f_0, x) = \beta_1 (f_1 - f_0) + \beta_3 x (f_1 - f_0)$$
(29)

Parametrelerin kestirilen değerleri yukarıdaki denklemde yerine konularak kestirilen log-odds değeri elde edilir ve varyans kestirimi aşağıdaki gibidir.

$$\hat{V}\{\ln[\hat{\Psi}(\mathrm{F} = f_0, \mathrm{F} = f_1, \mathrm{X} = \mathrm{x})]\} = \hat{V}(\hat{\beta}_1)(f_1 - f_0)^2 + \hat{V}(\hat{\beta}_3)[x(f_1 - f_0)]^2 + 2cov(\hat{\beta}_1, \hat{\beta}_3) x (f_1 - f_0)^2 \tag{30}$$

Çoğu istatistik yazılımı lojistik regresyon çözümlemesinde modele kestirilen parametrelerin varyans ve kovaryansı kestirimlerini vermektedir. Elde edilen kestirimler yukarıdaki eşitlikte yerine konarak kestirilen odds oranı için varyans hesaplanır. Kestirilen odds oranı için güven aralığı aşağıdaki eşitlikte verilmiştir.

$$\exp([\hat{\beta}_1(f_1 - f_0) + \hat{\beta}_3 x(f_1 - f_0)] \pm z_{1-\alpha/2} S\{\ln\left[\widehat{\Psi}(\mathrm{F} = f_0, \mathrm{F} = f_1, \mathrm{X} = \mathrm{x})\right]\} \quad (31)$$

Verilen bu eşitlikte standart hata (S), varyans kestiriminin kareköküdür.

4. UYGULAMA (SPSS)

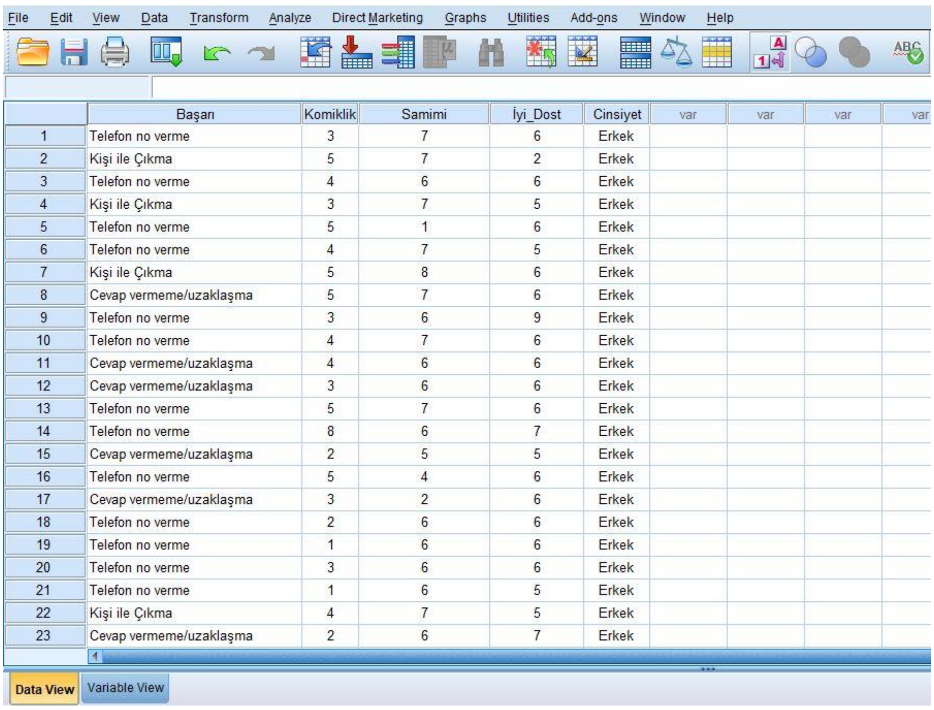
Bu tez de istatistik paket programlarından biri olan “SPSS” kullanılmıştır. Bu program İstatistiksel hesaplamalar ve grafikler için bir ortamdır. Etkin veri işleme ve saklama özelliğine sahiptir.

348 Erkek ve 672 kadın bireyle bir araştırma yapılıyor. Her birey karşı cinsle bir süre konuşuyorlar. Konuşmanın içerikleri; komiklik, samimi ve iyi_dost olmak üzere üç kategoriye ayrılıyor. Konuşmadan sonra bu üç bağımsız değişkene göre 0 ile 10 arasında bir puan veriliyor. (Örneğin komiklik kategorisinde 0 puan konuşmada hiç komiklik yok; 10 puan konuşma tamamen komik demektir.)

Yapılan konuşmalar sonucunda başarı başlığı altında bireyler arasında üç durum oluşuyor; kişi ile beraber çıkma, telefon numara verme, cevap vermeme/uzaklaşma.

Veri setimizi özetlersek; başarı değişkenimiz bağımlı bir değişken (sonuç değişkeni) ve üç kategoriden oluşmaktadır. Komiklik, samimi, iyi_dost ve cinsiyet değişkenlerimiz bağımsız değişkenlerdir (açıklayıcı değişkenler).

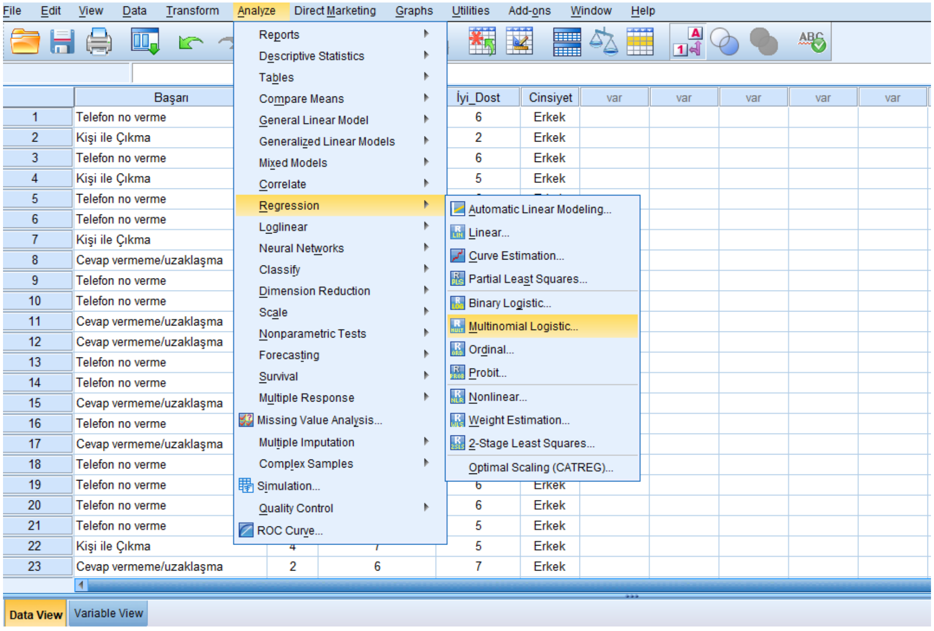
Burada bağımlı değişkenimiz olan başarı değişkenimiz üç kategoriden oluştuğu için yapacağımız analiz çok kategorili lojistik regresyon analizidir;



	Başarı	Komiklik	Samimi	İyi_Dost	Cinsiyet	var	var	var	var
1	Telefon no verme	3	7	6	Erkek				
2	Kişi ile Çıkma	5	7	2	Erkek				
3	Telefon no verme	4	6	6	Erkek				
4	Kişi ile Çıkma	3	7	5	Erkek				
5	Telefon no verme	5	1	6	Erkek				
6	Telefon no verme	4	7	5	Erkek				
7	Kişi ile Çıkma	5	8	6	Erkek				
8	Cevap vermeme/uzaklaşma	5	7	6	Erkek				
9	Telefon no verme	3	6	9	Erkek				
10	Telefon no verme	4	7	6	Erkek				
11	Cevap vermeme/uzaklaşma	4	6	6	Erkek				
12	Cevap vermeme/uzaklaşma	3	6	6	Erkek				
13	Telefon no verme	5	7	6	Erkek				
14	Telefon no verme	8	6	7	Erkek				
15	Cevap vermeme/uzaklaşma	2	5	5	Erkek				
16	Telefon no verme	5	4	6	Erkek				
17	Cevap vermeme/uzaklaşma	3	2	6	Erkek				
18	Telefon no verme	2	6	6	Erkek				
19	Telefon no verme	1	6	6	Erkek				
20	Telefon no verme	3	6	6	Erkek				
21	Telefon no verme	1	6	5	Erkek				
22	Kişi ile Çıkma	4	7	5	Erkek				
23	Cevap vermeme/uzaklaşma	2	6	7	Erkek				

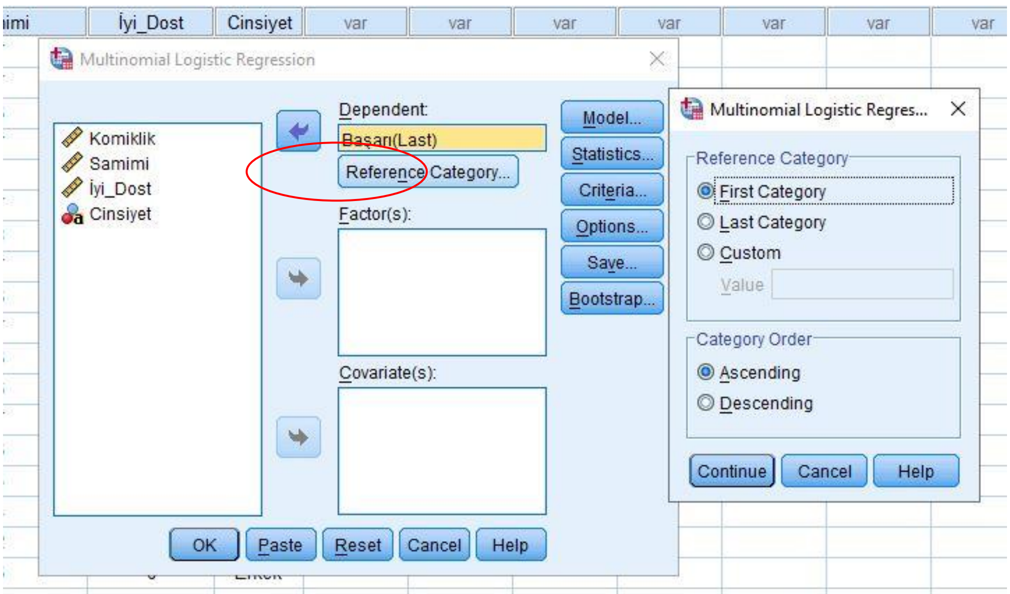
Şekil 1. Veri Seti

Uygulama verilerimiz görselde görüldüğü gibidir. Toplamda 1020 adet verimiz vardır.



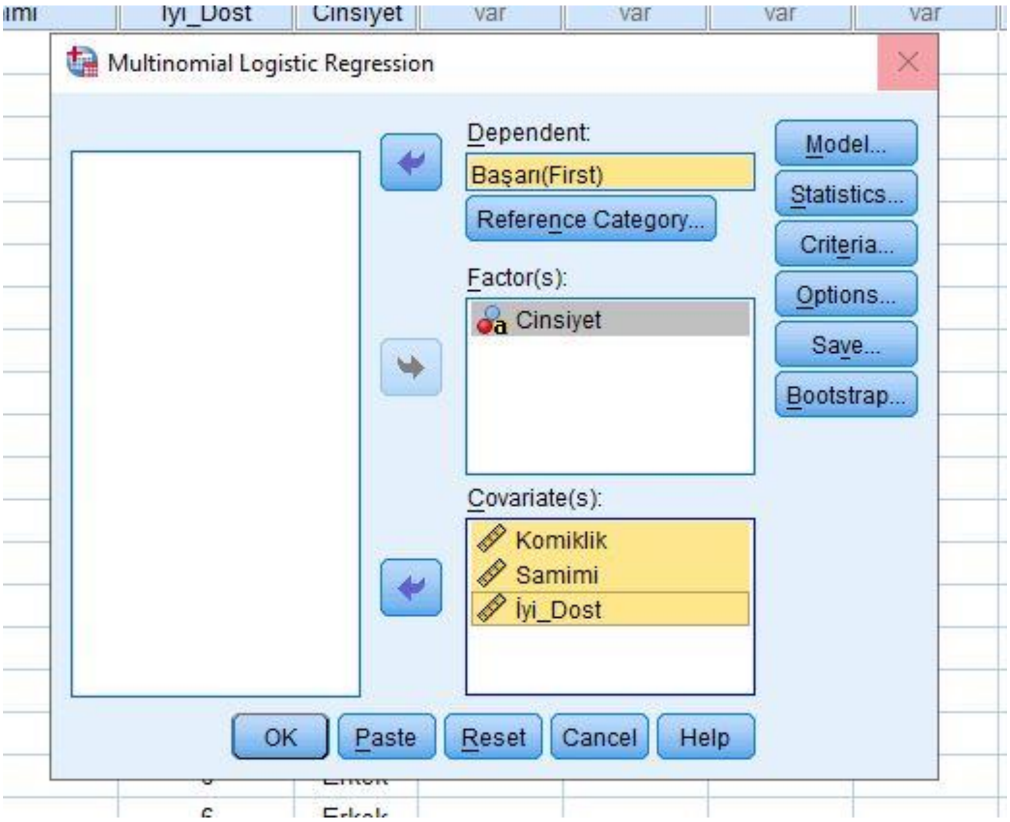
Şekil 2. Regresyon analizi menüsü

Analiz için menüden Analyze > Regression > Multinominal Lojistic kısmına geliyoruz.



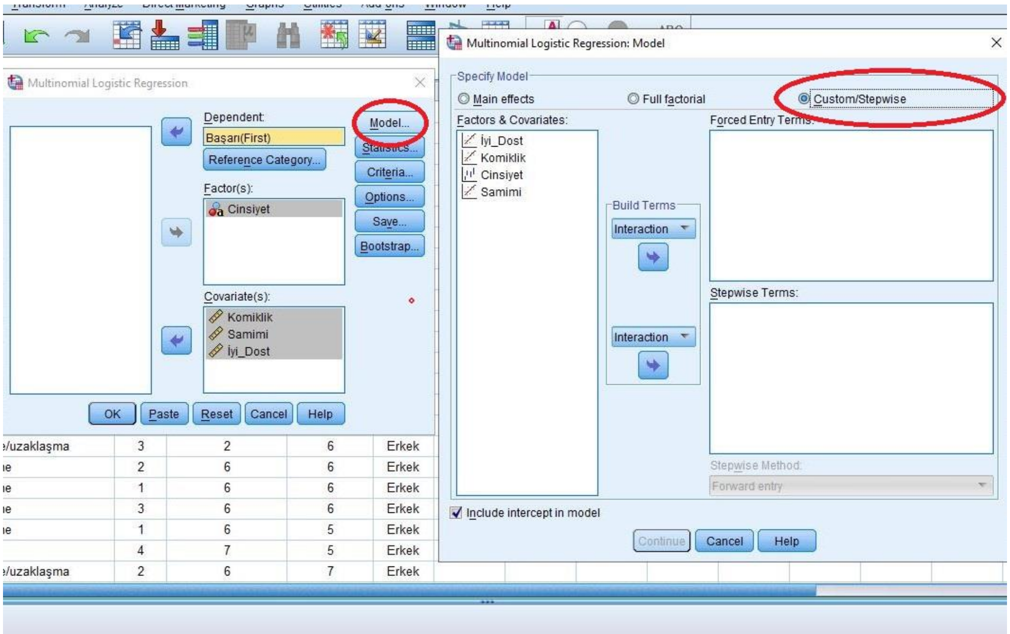
Şekil 3. Referans Kategori Belirleme

Açılan pencerede bağımlı değişkenimiz olan Başarı değişkenini Dependent kısmına atıyoruz. Burada iki kategorili lojistik regresyon analizinden farklı olarak Reference Category kısmından referans kategoriyi belirliyoruz. Biz First Category'i seçerek ilk kategoriyi referans kategori olarak aldık.



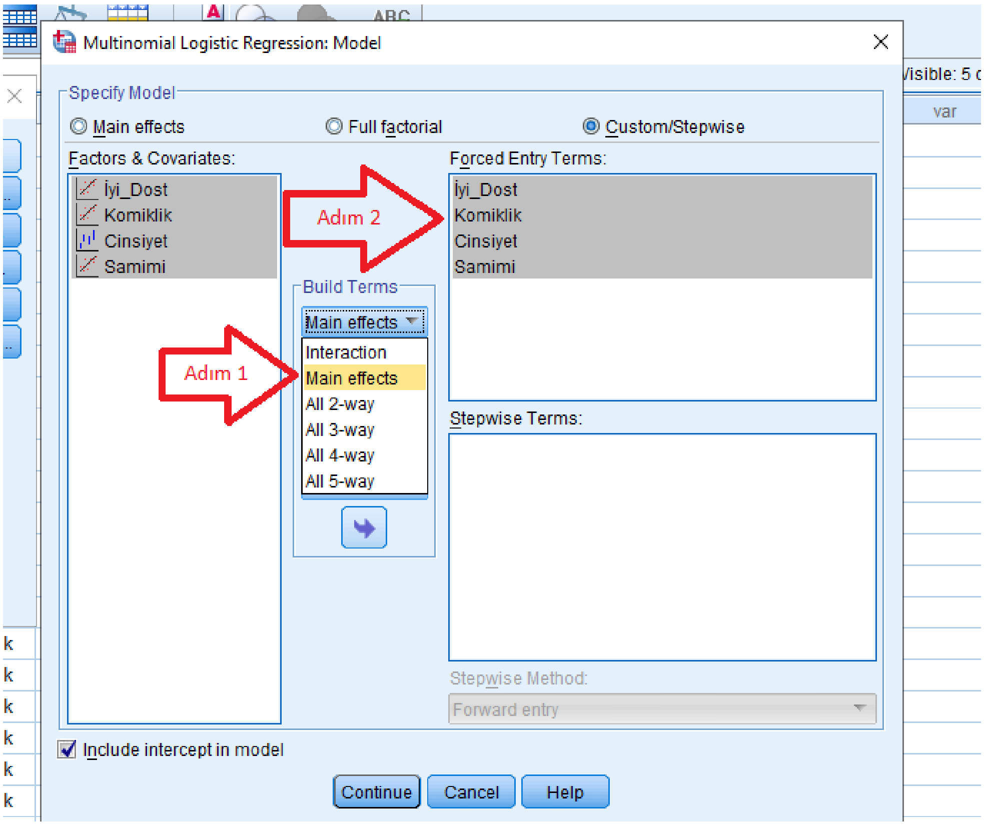
Şekil 4. Bağımlı bağımsız değişken ve kategorik değişken atama

Sonra gruplayıcı değişken olan Cinsiyeti Factor(s) kısmına ve Komiklik, Samimi, İyi_Dost bağımsız değişkenlerini Covariate(s) kısmına atıyoruz.



Şekil 4. Custom/Stepwise seçeneğini seçiyoruz.

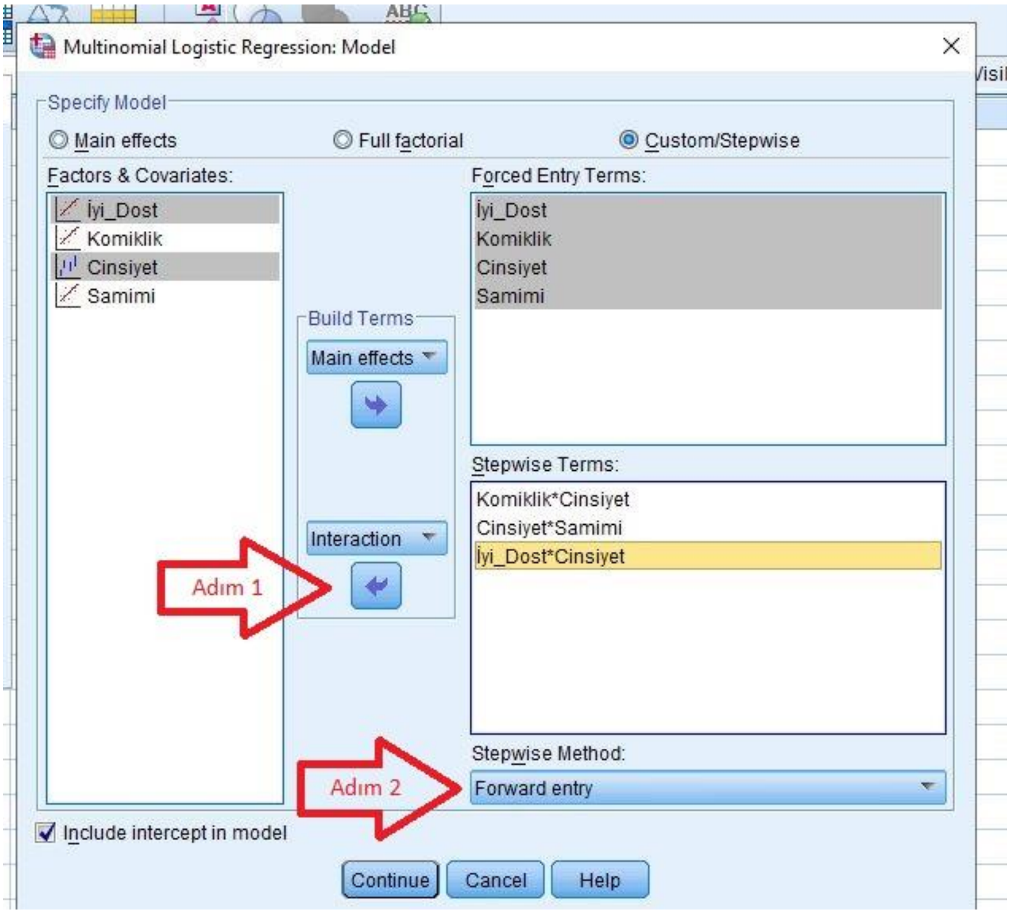
Etkileşim terimlerini belirlemek için model butonuna tıklıyoruz. Açılan pencerede Custom/Stepwise seçeneğini seçiyoruz.



Şekil 5. Multinominal Lojistic Regression: Model

Adım1: Build Terms kısmından Main effects (ana etkileri) seçiyoruz.

Adım2: İyi_Dost, Komiklik, Cinsiyet, Samimi bağımsız değişkenlerimizi ana etkil olarak Forced Entry Terms kısmına atıyoruz.

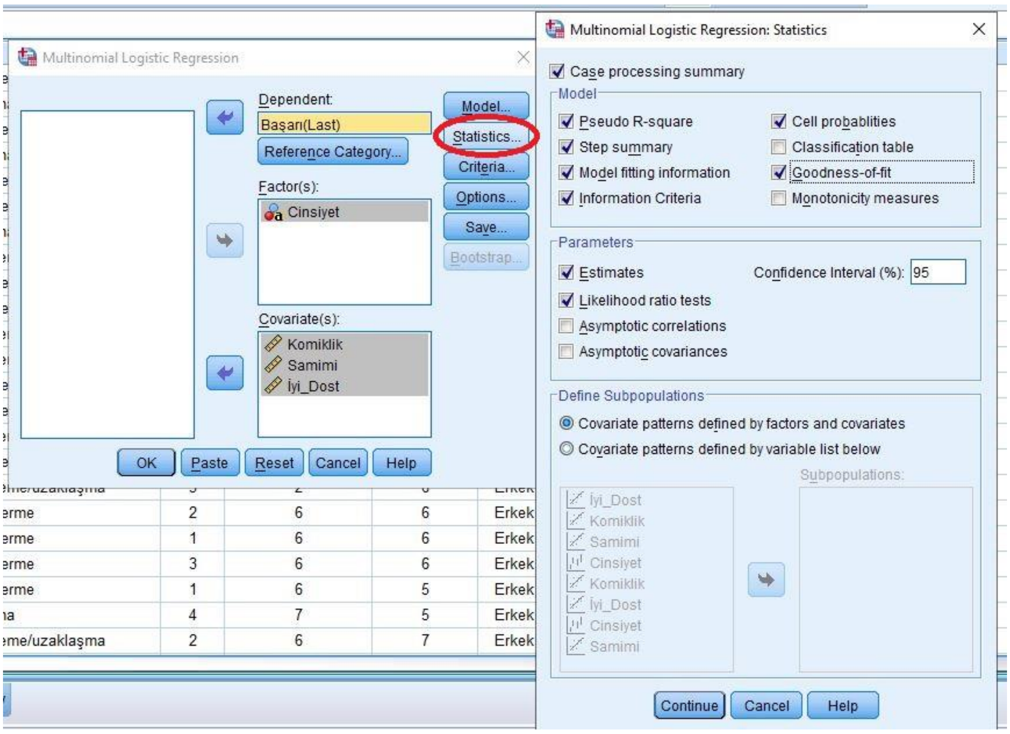


Şekil 6. Multinominal Lojistic Regression: Model

Sonra alt kısımda etkileşim terimlerini atıyoruz;

Adım1: Biz burada cinsiyet değişkeninin diğer değişkenler ile olan ikili etkilerini inceleyeceğiz. O yüzden cinsiyet değişkeniyle diğer üç değişkeni ikili olarak Stepwise Terms kısmına attık.

Adım2: Analiz metodu olarak en çok tercih edilen Forward entry metodunu seçiyoruz.



Şekil 7. Multinominal Lojistic Regression: Statistics

Model kısmındaki ayarlarımızı bitirdikten sonra Statistics kısmına geliyoruz. Sonra yukarıdaki görselde görüldüğü gibi analiz için gerekli olan bazı seçenekleri işaretliyoruz;

Pseudo R-square: Bu seçenek modelin R karesini verir.

Step summary: Adımsal prosedüre sahip olduğumuzdan adımların özetini verir.

Model fitting information: Modelin uygunluğu bilgisini verir.

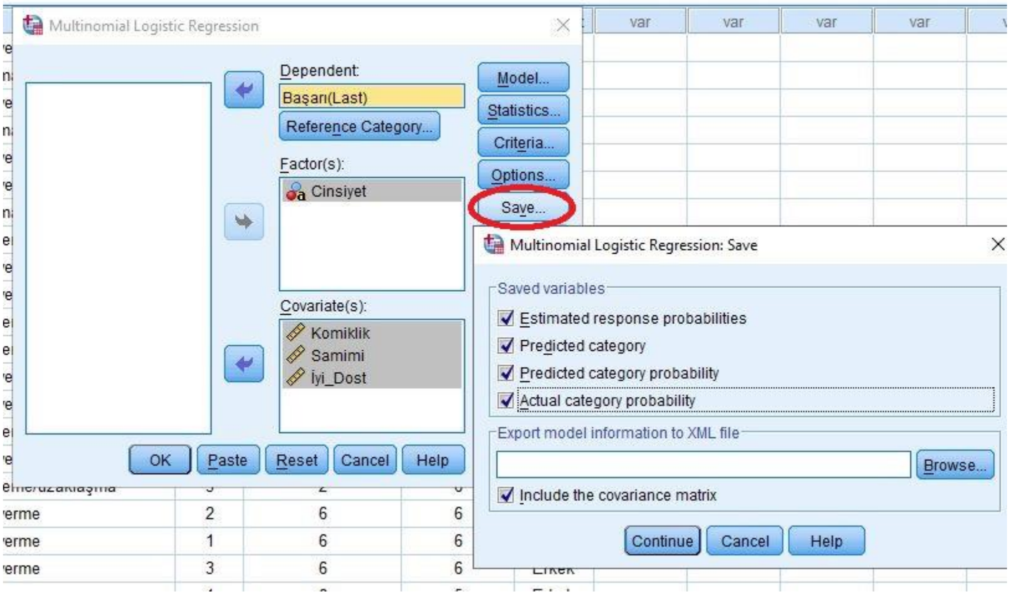
Information Criteria: Modelleri karşılaştırmak için kullanılır.

Cell probablities: Gözlenen ve beklenen frekansların tablosunu verir.

Goodness-of-fit: Modelin uygunluğu bilgisini verir.

Estimates: Bu seçenek çok önemlidir çünkü modeldeki tahminlerin beta değerlerini, test istatistiklerini ve güven aralıklarını verir.

Likelihood ratio tests: Bağımsız değişkenlerin tek tek önemliliği hakkında bilgi verir.



Şekil 8. Multinominal Lojistic Regression

Son olarak Save kısmından Estimatied response probilities, Predicted category, Predicted category probability ve Actual category probability seçeneklerini seçiyoruz. Bu seçenekler bize tahmin edilen olasılıkları ve tahmin edilen grup üyelerini seçer ve kaydeder. Bütün bu ayarlamaları yaptıktan sonra tamam deyip analiz sonuç tablolarını elde ediyoruz.

4.1. TABLOLAR VE YORUMLANMASI

Tablo 1. Case Processing Summary

Case Processing Summary

		N	Marginal Percentage
Başarı	Telefon no verme	485	47,5%
	Kişi ile Çıkma	135	13,2%
	Cevap vermeme/uzaklaşma	400	39,2%
Cinsiyet	Kadın	672	65,9%
	Erkek	348	34,1%
Valid		1020	100,0%
Missing		0	
Total		1020	
Subpopulation		314[a]	

a. The dependent variable has only one value observed in 212 (67,5%) subpopulations.

Bu tabloda Valid gözlenen değerlerdir. Missing eksik gözlem değerleridir.

Tablo 2. Adım Özeti Tablosu

Step Summary

Model	Action	Effect(s)	Model Fitting Criteria			Effect Selection Tests		
			AIC	BIC	-2 Log Likelihood	Chi-Square[a]	df	Sig.
0	Entered	Intercept, İyi_Dost, Komiklik, Cinsiyet, Samimi	937,572	986,848	917,572	.		
1	Entered	Cinsiyet * Komiklik	908,451	967,582	884,451	33,121	2	,000
2	Entered	Cinsiyet * Samimi	899,002	967,987	871,002	13,450	2	,001

Burada model 0 bize bağımsız değişkenlerin modele girmesiyle oluşan modelin bilgisini verir. Model 1 cinsiyet*komiklik etkileşim değişkeninin modele girmesiyle oluşan modelin bilgisini verir. Model 2 cinsiyet*samimi etkileşim değişkeninin modele girmesiyle oluşan modelin bilgisini verir.

Dikkat edersek AIC (Akaiki bilgi kriteri), BIC (bayesian bilgi kriteri) ve -2 Log Likelihood değerleri gitgide küçülüyor. Bu da bize etkileşim değişkenlerinin modele girmesiyle anlamlı bir farkın oluşacağını gösterir.

Ayrıca Sig.(P-Value) değerlerinin 0,05 alfa değerinden küçük olması yine etkileşim değişkenlerinin modele girmesiyle anlamlı bir farkın oluşacağının göstergesidir.

Tablo 3. Model Uygunluk Bilgisi

Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1153,526	1163,382	1149,526			
Final	899,002	967,987	871,002	278,525	12	,000

Burada -2 Log Likelihood değerine baktığımızda ilk başta hiçbir değişken olmadığından modelin 1149,526 değeri açıklanamıyor. Ama en son adımda yani bütün değişkenler modele girdiği zaman bu açıklanamayan değer 871,002'ye düşüyor. Aradaki fark 278,525 dir. Sig.(P)değeri 0,05'ten küçük olduğu için en son adımda oluşan model istatistiksel olarak anlamlı demektir. Ki-kare değeri de 278,525 olarak hesaplanmıştır.

Tablo 4. Uyum İyiliği Tablosu

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	886,616	614	,000
Deviance	617,481	614	,453

Pearson ve Deviance değerleri aynı durumu ifade eder. Modelin veriye uygun olabilmesi için; Pearson değeri <0,05 veya Deviance değeri >0,05 olmalıdır. Bizim verimiz modele uygundur.

Tablo 5. R-kare Tablosu

Pseudo R-Square

Cox and Snell	,239
Nagelkerke	,277
McFadden	,138

Lojistik regresyonda R-kare değerleri doğrusal regresyondan farklı olarak küçük çıkma eğilimindedir. Bazı istatistikçiler burada bu değerlerin bir ölçü olmadığını ifade etmektedirler.

Tablo 6. Likelihood Oran Testlerinin Tablosu

Likelihood Ratio Tests

Effect	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC of Reduced Model	BIC of Reduced Model	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	899,002	967,987	871,002[a]	,000	0	.
İyi_Dost	901,324	960,454	877,324	6,322	2	,042
Komiklik	899,002	967,987	871,002[a]	,000	0	.
Cinsiyet	913,540	972,671	889,540	18,538	2	,000
Samimi	899,002	967,987	871,002[a]	,000	0	.
Cinsiyet * Komiklik	930,810	989,941	906,810	35,808	2	,000
Cinsiyet * Samimi	908,451	967,582	884,451	13,450	2	,001

Bağımsız değişkenlerin tek tek anlamlılığını gösterir. Sig.(P) değeri 0,05'ten küçük olan bağımsız değişkenlerin analiz için önemli olduğunu anlıyoruz.

Tablo 7. Parametre Tahminleri Tablosu

Parameter Estimates

Başarı[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	95% Confidence Interval for Exp(B) Upper Bound
Telefon no verme	Intercept	-1,783	,670	7,087	1	,008			
	İyi_Dost	,132	,054	6,022	1	,014	1,141	1,027	1,268
	Komiklik	,139	,110	1,602	1	,206	1,150	,926	1,427
	[Cinsiyet=Female]	-1,646	,796	4,274	1	,039	,193	,040	,918
	[Cinsiyet=Male]	0[b]	.	.	0	.	.	.	.
	Samimi	,276	,089	9,589	1	,002	1,318	1,107	1,570
	[Cinsiyet=Female] * Komiklik	,492	,140	12,374	1	,000	1,636	1,244	2,153
	[Cinsiyet=Male] * Komiklik	0[b]	.	.	0	.	.	.	.
	[Cinsiyet=Female] * Samimi	-,348	,106	10,824	1	,001	,706	,574	,869
	[Cinsiyet=Male] * Samimi	0[b]	.	.	0	.	.	.	.
Kişi ile Çıkma	Intercept	-4,286	,941	20,731	1	,000			
	İyi_Dost	,130	,084	2,423	1	,120	1,139	,967	1,341
	Komiklik	,318	,125	6,459	1	,011	1,375	1,076	1,758
	[Cinsiyet=Female]	-5,626	1,329	17,934	1	,000	,004	,000	,049
	[Cinsiyet=Male]	0[b]	.	.	0	.	.	.	.
	Samimi	,417	,122	11,683	1	,001	1,518	1,195	1,928
	[Cinsiyet=Female] * Komiklik	1,172	,199	34,627	1	,000	3,230	2,186	4,773
	[Cinsiyet=Male] * Komiklik	0[b]	.	.	0	.	.	.	.
	[Cinsiyet=Female] * Samimi	-,477	,163	8,505	1	,004	,621	,451	,855
	[Cinsiyet=Male] * Samimi	0[b]	.	.	0	.	.	.	.

a. The reference category is: Cevap vermeme/uzaklaşma.

b. This parameter is set to zero because it is redundant.

Bağımsız değişkenlerin etkisinin ne kadar olduğunu, modele ne kadar etki ettiğini gösteren tablodur. Tabloda Sig.(P) değeri 0,05'ten küçük olan değişkenler model için anlamlıdır.

Exp(B) değeri bize başarı kategorilerine yaklaşma(başarma) ihtimalini gösterir.

Örneğin telefon no verme kategorisinde;

İyi_Dost değişkeninin Sig. değeri 0,05'ten küçüktür. Exp(B) değeri 1,141'dir. Yani kişi İyi_Dost konuşmasında bir birimlik artış gösterirse telefon no verme olasılığını 1,027 kat artar.

Komiklik değişkenine baktığımız zaman Sig. değeri(0,206) 0,05'ten büyüktür. Yani komiklik değişkeni model için anlamlı değildir.

[Cinsiyet=Female] değişkeninin Sig. değeri 0,05'ten küçüktür. Exp(B) değeri 0,193'tür. Bu değer 1'den küçük olduğu için kadınların telefon no verme olasılığı daha düşüktür denir.

[Cinsiyet=Female]*Komiklik etkileşim değişkeninin Sig. değeri 0,05'ten küçüktür. Exp(B) değeri 1,636'dır. Bu değişken bize komiklik arttıkça kadınların telefon no verme olasılığının daha yüksek olduğunu gösteriyor.

[Cinsiyet=Female]*Samimi etkileşim değişkeninin Sig. değeri 0,05'ten küçüktür. Exp(B) değeri 0,706'dır. Bu değer 1'den küçük olduğu için Samimi sohbet arttıkça kadınların telefon no verme olasılığı daha düşüktür sonucuna varırız.

Cevap vermeme/uzaklaşma kategorisi referans kategori olarak kabul edildiği için burada göremiyoruz.

Kişi ile çıkma kategorisini de aynı şekilde yorumluyoruz.

Tablo 8. Sonuç Tablosu (Telefon No Verme*Cevap Vermeme)

	B	Standart Hata	Odds Oranı(Exp)	Güven Aralığı
Sabit	-1.78**	0.67		
İyi Arkadaşlık	0.13*	0.05	1.14	(1.03-1.27)
Komiklik	0.14	0.11	1.15	(0.93-1.43)
Cinsiyet	-1.65*	0.80	0.19	(0.04-0.92)
Samimi	0.28**	0.09	1.32	(1.11-1.57)
Cinsiyet*Komiklik	0.49***	0.14	1.64	(1.24-2.15)
Cinsiyet*Samimi	-0.35**	0.11	0.71	(0.57-0.87)

*: $p<0,05$; **: $p<0.01$; ***: $p<0.001$

B: Değişkenlerin katsayıları

Tablo 9. Sonuç Tablosu (Kişi İle Çıkma*Cevap Vermeme)

	B	Standart Hata	Odds Oranı(Exp)	Güven Aralığı
Sabit	-4.29***	0.94		
İyi Arkadaşlık	0.13	0.08	1.14	(0.97-1.34)
Komiklik	0.32*	0.13	1.38	(1.08-1.76)
Cinsiyet	-5.63***	1.33	0	(0-0.05)
Samimi	0.42**	0.12	1.52	(1.20-1.93)
Cinsiyet*Komiklik	1.17***	0.20	3.23	(2.19-4.77)
Cinsiyet*Samimi	-0.48**	0.16	0.62	(0.45-0.86)

*: $p<0,05$; **: $p<0.01$; ***: $p<0.001$

5. BULGULAR VE SONUÇLAR

Regresyon analizi yapabilmek için bağımlı değikenimiz kategorik olmalıdır. Bu değişken ikiden fazla kategoriye sahip olduğunda çok değişkenli (multinominal) lojistik regresyon analizi uyguluyoruz. Analiz sonucunda bağımsız değişkenler arasındaki ilişkiyi çeşitli amaçlar doğrultusunda incelemiş, aralarındaki ilişkiyi anlamış oluyoruz.

Çok kategorili lojistik regresyon analizi yaparken kategorilerden biri referans alınır. Referans alınan kategoriye göre diğer kategoriler incelenir. Yani k tane kategorik değişkenimiz varsa ; k-1 karşılaştırma yapılır.

Karşılaştırma sonucunda veri uygun bir modele sahip olmuş olur. Yani minimum değişken ile en iyi uyuma sahip olacak biçimde tanımlayabilen, kabul edilebilir model kurulmuş olur.

6. KAYNAKLAR

Alpar, R. (2011). Uygulamalı çok değişkenli istatistiksel yöntemler. Ankara: Detay Yayıncılık.

Atasoy, D. (2001). Lojistik regresyon analizinin incelenmesi ve bir uygulaması. (Yüksek lisans tezi, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas).

Barak, A., Karahan, S. ve Saraçbaşı, O. (2005, Mayıs). Ordinal lojistik regresyon modelleri. 4. İstatistik Kongresi'nde sunulan bildiri, Belek-Antalya.

Cook D., Dixon P., Duckworth W. M., Kaiser M. S., Koehler K., Meeker W. Q. & Stephenson W. R. (2001).Binary response and logistic regression analysis, in Part of the Iowa State University NSF/ILI project. Beyond Traditional Statistical Methods www.public.iastate.edu/ ~stat415/stephenson/stat415_chapter3.pdf. adresinden edinilmiştir.

Cox, R.,(1983). Some remarks on overdispersion. Biometrika,70, 269-274. http://www.jstor.org/stable/2335966 adresinden edinilmiştir.

Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010). Sosyal bilimler için çok değişkenli istatistik- spss ve lisrel uygulamaları. Ankara: Pegem Akademi Yayınları.

Çokluk. Ö. (2010). Lojistik regresyon analizi: kavram ve uygulama. Kuram ve Uygulamada Eğitim Bilimleri,10(3),1357-1407. http://www.edam.com.tr/kuyeb/pdf/tr/3e2b1f84ce847e4fef09b68db9b1a420kFULL.pdf adresinden edinilmiştir.

Dean, C. B. (1992). Testing for overdispersion in poisson and binomial regression models. Journal of American Statistical Association, 87(418), 451-457. http://www.jstor.org/stable/2290276 adresinden edinilmiştir.

Elhan, A. H. (1997). Lojistik regresyon analizinin incelenmesi ve tıpta bir uygulaması (Yüksek Lisans Tezi Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara).

https://en.wikipedia.org/wiki/Logistic_regression