



Kategorik Verilerin Analizinde Logaritmik Doğrusal Modellerin Kullanımı: İntihar Olasılığı Verileri Üzerine Bir Uygulama¹

Esra TOPALOĞLU*, Aliye ATAY**

ÖZ

Sosyal bilim alanlarında yapılan araştırmalar sonucu elde edilen veriler genellikle kategorik veriler olup, parametrik analiz varsayımları sağlanmadığından parametrik olmayan yöntemler kullanılmaktadır. Kategorik değişkenler arasındaki ikili ilişkilerin test edilmesinde Pearson test istatistiği yeterli olmakta ancak ikiden fazla değişken olması durumunda bu istatistik kullanılamamaktadır. Bu nedenle çalışmada iki veya daha fazla değişken arasındaki ilişkilerin test edilmesine olanak sağlayan Logaritmik Doğrusal Modeller (Log-Linear) kullanılarak cinsiyet, sosyoekonomik düzey ve intihar olasılığından oluşan değişkenlerin birbirleri ile olan etkileşim yapısı 15-25 yaş grubu arasındaki öğrenciler için incelenmiştir. Elde edilen sonuçlara göre erkeklerin intihar olasılığı yüksek grupta olma oranı kadınlara göre daha fazladır ve gruplar arasındaki bu fark istatistiksel olarak anlamlıdır.

Anahtar Kelimeler: Kategorik Veri, Logaritmik Doğrusal Modeller, İntihar Olasılığı

JEL Sınıflandırması: C10, C25

Use of Logarithmic Linear Models in Analysis of Categorical Data: An Application on Suicide Risk Data

ABSTRACT

The data obtained as a result of researches in social sciences are generally categorical, and nonparametric methods are used since the assumptions of parametric analysis are not provided. The Pearson test statistic is sufficient to test the binary relationships between categorical variables, but this statistic cannot be used if there are more than two variables. Therefore, in the study, using the Logarithmic Linear (Log-Linear) Models that allow testing the relationships between two or more variables, the interaction structure of variables consisting of gender, socioeconomic level, and suicide probability were examined for students between the ages of 15-25. Based on the results, it was found that, the rate of men being in the high-risk suicide group is higher than the rate of women in the same group and this difference between the groups is statistically significant.

Keywords: Categorical Data, Log-Linear Models, Suicide Risk

JEL Classification: C10, C25

Geliş Tarihi / Received: 11.02.2020 Kabul Tarihi / Accepted: 14.05.2020

¹ Bu çalışma, Esra TOPALOĞLU'nun "Logaritmik Doğrusal Modeller ve Bir Uygulaması" adlı tezinden uyarlanarak oluşturulmuştur.

* Bilim Uzmanı, Akdeniz Üniversitesi, SBE, Ekonometri Bölümü, topalogluesra3@gmail.com, ORCID: 0000-0001-9374-5404.

** Dr. Öğr. Üyesi, Türk Hava Kurumu Üniversitesi, İşletme Fakültesi, Havacılık Yönetimi Bölümü, aatay@thk.edu.tr, ORCID: 0000-0003-0812-6575.

1. GİRİŞ

Bir araştırmanın istatistiksel değerlendirme süreci yapılırken verilere uygun istatistiksel tekniklerin kullanılması önemlidir. Bu anlamda ilgilenilen veriler için hangi ölçeğin kullanılacağı ve ele alınan değişkenler arasındaki ölçüm düzeylerine göre kullanılacak istatistiksel teknikler farklılık göstermektedir.

Sosyal bilim alanlarında yapılan araştırmalar sonucu elde edilen veriler genellikle kategorik veriler olup, parametrik analiz varsayımları sağlanmadığından parametrik olmayan yöntemler kullanılmaktadır. Ancak kategorik verilerin analizinde kullanılan yöntemlerin tarihsel gelişim süreci çok eskiye dayanmamaktadır. 1960'lı yıllara kadar kategorik verilere uygulanacak istatistiksel metotların gelişimi nicel verilere uygulanan metotların gelişiminden geride kalmıştır. Bu alanda 1900'lü yıllarda Karl Pearson tarafından etkili çalışmalar yapılmıştır. Logaritmik doğrusal model uygulamalarını ele alan çalışmalar 1960'lı yıllardan sonra gelişmeye başlamış, Goodman (1970), Haberman (1974), Agresti (1990) gibi bilim insanlarının çalışmaları ile kategorik verilerin analizinde Logaritmik Doğrusal (Log-Linear) Model uygulamalarının kullanımı önem kazanmıştır.

Logaritmik doğrusal modellerde gözlem sonucu elde edilen veriler çeşitli çapraz tablolar halinde düzenlenerek analiz edilmektedir. İki kategorik değişkenin ele alındığı iki yönlü çapraz tablolarda sadece bağımsızlık test edilmekte iken, üç veya daha fazla sayıda değişkenin oluşturduğu çapraz tabloların analizinde temel ve etkileşim etkilerine ait hipotezler de test edilmektedir. Bu nedenle logaritmik doğrusal model analizi kesikli çok değişkenli analizler arasında önemli bir yere sahiptir. Logaritmik doğrusal modellerde incelenen tüm değişkenler yanıt değişkenleri olarak ele alınmaktadır. Diğer bir deyişle, değişkenler arasında açıklanan ve açıklayıcı değişken ayrımı yapılmamaktadır. Bu nedenle, logaritmik doğrusal modeller yalnızca değişkenler arasındaki ilişki yapısını ortaya koymaktadır (Jeansonne, 2017).

Literatürde çok değişkenli kategorik verilerin analizinde kullanılan logaritmik modellerle ilgili çeşitli çalışmalar bulunmaktadır. Yılmaz ve Şıklar (2002), intihar olayları ile ilgili çok değişkenli kategorik verileri logaritmik doğrusal modeller yardımı ile incelemiş ve araştırma sonucunda iki ve üç değişkenli etkileşimler istatistiksel olarak anlamlı bulunmuştur. Bülbül (2006), çocuk işgücü değişkenleri arasındaki etkileşimleri üç yönlü çapraz tablolar halinde düzenleyerek logaritmik doğrusal analiz ile incelemiştir. Filiz (2007), üç yönlü logaritmik doğrusal modeller ile üniversite öğrencilerinde sigara, alkol ve nargile içme sıklığını belirlemek ve bunları etkileyen risk faktörlerini incelemek amacıyla çalışma yapmıştır. Mete ve Ünsal (2010), logaritmik doğrusal modeller ile göç istatistikleri üzerine bir uygulama yapmıştır. Kaşıkır (2012), sigara içen lise öğrencilerinin sigara içmelerinde etkili olan nedenleri logaritmik doğrusal modeller ile ortaya koymaya çalışmıştır. Erdem (2014), logaritmik doğrusal modeller yardımı ile çeşitli program türlerinin izlenme durumu, en çok izlenen televizyon kanalları ve bireylere ilişkin çeşitli demografik değişkenler arasındaki ilişkiyi incelemiştir. Yurt Öncel ve Erdugan (2015), logaritmik doğrusal modeller ile sigara bağımlılığı üzerine bir uygulama yapmıştır. Erdugan ve Türkan (2017), iş kazalarını, üç yönlü çapraz tablolar halinde düzenleyerek logaritmik doğrusal analiz yöntemi aracılığı ile incelemiştir. Köleoğlu (2018), uluslararası öğrencilerin sosyal uyumlarını logaritmik doğrusal modeller kullanarak analiz etmiştir. Kandemir ve Şimşek (2019), geçici iş göremezlik süresi, cinsiyet, çalışılan ortam ve iş kazası türleri etkileşimlerini logaritmik doğrusal modeller ile analiz etmiştir.

Bu çalışmada, Türk Dil Kurumu (TDK) tarafından bir kimsenin toplumsal ve ruhsal nedenlerin etkisi ile kendi hayatına son vermesi; Dünya Sağlık Örgütü (DSÖ) tarafından ise “kişinin amacının bilincinde ve değişik derecelerde ölümcül amaçlı olarak kendine zarar vermesi” olarak tanımladığı intihar olgusu üzerinde durulmuştur. Dünya Sağlık Örgütü'nün yayımladığı bir rapora göre, dünyada her 40 saniyede bir kişi intihar etmekte; 15-29 yaş arasındaki gençlerde yol kazalarının ardından en büyük ikinci ölüm nedeninin intihar olduğu

gösterilmektedir. Ayrıca Türkiye İstatistik Kurumu (TÜİK) tarafından yayımlanan verilere göre 2018 yılında 3 bin 161 kişinin intihar ettiği; yani her gün ortalama 9 kişinin hayatına son verdiği ortaya konulmaktadır.

Hem dünyada hem de ülkemizde 15-24 yaş grubu arasındaki gençler intihar açısından riskli grup olarak değerlendirilmektedirler (Atay, Eren ve Gündoğar, 2012; Sayıl ve ark., 1998; Sayıl, Devrimci Özgüven 2002; Simons ve Murphy, 1985). Bu nedenle bu yaş grubunda bulunan gençlerin intihar olasılığının değerlendirilmesinin artan intihar oranlarına önlem alma açısından önemli olduğu düşünülmektedir. Bu amaçla Hisli Şahin ve Durak Batıgün (2009), tarafından gerçekleştirilen “Lise ve Üniversite Öğrencilerinde İntihar Riskini Belirlemeye Yönelik Bir Modelin Sınanması” çalışmasına ait verileri kullanılarak Ankara ilinde çeşitli lise ve üniversitelerde okuyan 15-25 yaş grubu arasındaki 2343 öğrenci için intihar olasılığının; bireyi intihara sürükleyebilecek riskler arasında düşünüldüğü cinsiyet, sosyoekonomik düzey gibi değişkenlerle olan etkileşim yapısı Logaritmik Doğrusal Modeller ile incelenmiştir.

2. YÖNTEM

Sınıflama (nominal) ve sıralama (ordinal) ölçüm düzeyine sahip veriler olarak nitelendiren kategorik verilerin analizinde değişkenler arasındaki ilişkilerin test edilmesinde çapraz tablolar (kontenjans tabloları, olumsuzluk tabloları vb.) kullanılmaktadır.

Çapraz tablolar değişken sayısı k olmak üzere iki yönlü, üç yönlü veya k yönlü çapraz tablolar olarak adlandırılmakta ve bu tablolardaki hücrelerde yer alan değerler sayım yolu ile belirlenmektedir. Dolayısıyla kategorik veri, frekans verileri yani sayım verileri olarak da ifade edilmekte ve değişken kategorilerinin veri kümesinde kaç kez tekrarlandığını göstermektedir.

Kategorik verilerin analizinde çapraz tablolar esas alınmakta ve ikili ilişkilerin incelenmesinde sıklıkla χ^2 analizi veya L^2 , olabilirlik oran istatistiği (Likelihood Ratio Statistics) kullanılmaktadır. Ancak, ikili, üçlü ve daha karmaşık yapıları ilişkilerin incelenmesinde geliştirilmiş logaritmik doğrusal modellerin bir uzantısı olan “Logaritmik Doğrusal Modeller” (Log-Linear Models) kullanılmaktadır. Logaritmik Doğrusal Modeller değişkenler arasındaki ikili etkileşimlerin yanı sıra üçlü ve daha fazla etkileşimi içerdiğinden uygulamada kullanımı önemli ve yaygındır.

2.1. Logaritmik Doğrusal Modeller

Kategorik verilerin analizinde sıklıkla kullanılan logaritmik doğrusal modeller, Poisson dağılımına sahip veri için geliştirilmiş doğrusal modellerin özel bir durumudur. Genellikle iki veya daha fazla değişkenin (çok yönlü çapraz tabloların) analizinde kullanıldığından “Çok-Yönlü Frekans Analizi” olarak da adlandırılmaktadır.

Logaritmik doğrusal modeller karmaşık yapıları çok yönlü tabloların analizine sistematik bir yaklaşım getirmektedir. İlgilenilen etkilerin büyüklüğünün tahmin edilmesine ve buna bağlı olarak incelenecek farklı etkilerin göreceli önemini belirlemesine olanak sağlamaktadır. Ayrıca iki kategorik değişkenin ele alındığı iki yönlü çapraz tablolarda sadece bağımsızlık test edilmekte iken, üç veya daha fazla sayıda değişkenin oluşturduğu çapraz tabloların analizinde temel ve etkileşim etkilerine ait hipotezler de test edilmektedir. Bu nedenle logaritmik doğrusal model analizi kesikli çok değişkenli analizler arasında önemli bir yere sahiptir.

Logaritmik doğrusal modellerde incelenen değişkenler arasında açıklanan ve açıklayıcı değişken ayrımı yapılmamakta olup bu modeller yalnızca değişkenler arasındaki ilişki yapısını ortaya çıkarmaktadır. Ancak bir ya da daha fazla değişken arasında açıklanan değişken ve

açıklayıcı değişken ayrımı yapılmak isteniyorsa logaritmik doğrusal modeller yerine logit ya da lojistik regresyon modeli kullanılmaktadır.

Uygun (1990: 290), iki ya da daha fazla değişken içeren çapraz tabloların analizinde logaritmik doğrusal modellerin kullanım amacını,

1. Değişkenlerin oluşturduğu bileşik dağılımı test etmek,
 2. Değişkenlerin birbirlerine bağımlı olup olmadığını test etmek,
 3. Değişkenler arasındaki ilişkiyi neden-sonuç ilişkisine dayandırmaksızın test etmek,
- olarak üç başlık altında toplamıştır.

Değişkenler arasındaki karmaşık ilişki yapısını ortaya çıkaran aynı anda birden fazla hipotezin test edilmesine olanak sağlayan logaritmik doğrusal modeller, dört veya daha fazla değişken içeren çok yönlü çapraz tabloların ($k \geq 4$) analizinde de kullanılmaktadır. Ancak değişken sayısı arttıkça bağımsızlık eşitlikleri de buna bağlı olarak artar ve etkileşim terimlerini açıklamak karmaşık, kolay yorumlanamayan bir hal almaktadır. Bu nedenle çalışmada, daha anlaşılabilir olması açısından model yapısı, parametre kestirimleri ve yorumları çalışmada kullanılan değişken sayısı itibari ile üç yönlü çapraz tablolar üzerinden ifade edilmiştir.

2.2. Üç Yönlü Çapraz Tablolarda Logaritmik Doğrusal Modeller

Üç değişkenin yer aldığı bir çapraz tablo için sırasıyla i , j ve k indisli A, B ve C değişkenlerinin arasındaki ilişkinin incelenmek istendiğini varsayalım.

i (sıra), j (sütun) ve k (tabaka)'yı ifade etmek üzere (i, j, k) gözesinin gözlenen frekansını n_{ijk} ile gösterelim. A değişkeninin i 'inci sıra, B değişkeninin j 'inci sütun ve C değişkeninin k 'inci tabakasına düşme olasılığını π_{ijk} ile ifade edelim. Bu durumda, n_{ijk} 'ların olasılık dağılımını A, B ve C değişkenlerinin ortak olasılık dağılımı verir.

Beklenen sıklıkları m_{ijk} ile göstermek üzere, A, B ve C kategorik değişkenleri için sırasıyla $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; $k = 1, 2, \dots, K$ düzeylerinin her bir kombinasyonu için oluşturulacak $I \times J \times K$ boyutlu bir çapraz tablonun oluşturacağı logaritmik doğrusal model aşağıdaki gibi tanımlanır:

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} .$$

Bu model “Doymuş Logaritmik Doğrusal Model” olarak adlandırılmaktadır. Bu model, aynı zamanda “Hiyerarşik Logaritmik Doğrusal Model” olarak ifade edilmekte ve $\log(m_{ijk})$ beklenen frekansların doğal logaritması olmak üzere, λ_0 genel ortalama, değişkenlere ait ana etkiler $\{\lambda_i^A, \lambda_j^B, \lambda_k^C\}$, bütün olası ikili etkileşim etkiler $\{\lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC}\}$ ve üçlü etkileşim etkiler $\{\lambda_{ijk}^{ABC}\}$ parametrelerini içermektedir.

Doymuş modele ait parametreler ve serbestlik dereceleri aşağıdaki tabloda yer almaktadır.

Tablo 2.1. Üç Yönlü Logaritmik Doğrusal Model Parametre ve Serbestlik Dereceleri

Parametreler	Serbestlik Dereceleri
λ_0	1
λ_i^A	$(I - 1)$
λ_j^B	$(J - 1)$
λ_k^C	$(K - 1)$
λ_{ij}^{AB}	$(I - 1)(J - 1)$
λ_{ik}^{AC}	$(I - 1)(K - 1)$
λ_{jk}^{BC}	$(J - 1)(K - 1)$
λ_{ijk}^{ABC}	$(I - 1)(J - 1)(K - 1)$

Üç yönlü çapraz tablolara ait toplam 9 adet logaritmik doğrusal model ise Tablo 2.2.'de verilmiştir (Andersen, 1990).

Tablo 2.2. Üç Yönlü Çapraz Tablolarda Olası Hiyerarşik Logaritmik Doğrusal Modeller

Model	Gösterim	Model
$M^{(0)}$	[A][B][C]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C$
$M^{(1)}$	[AB][C]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$
$M^{(2)}$	[AC][B]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC}$
$M^{(3)}$	[BC][A]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$
$M^{(4)}$	[AB][BC]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC}$
$M^{(5)}$	[AB][AC]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$
$M^{(6)}$	[AC][BC]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$
$M^{(7)}$	[AB][AC][BC]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$
$M^{(8)}$	[ABC]	$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$

Tablo 2.2. incelendiğinde $M^{(8)}$ modeli ana etkileri ($\lambda_i^A, \lambda_j^B, \lambda_k^C$), tüm ikili etkileşimleri ($\lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC}$) ve üçlü etkileşimleri (λ_{ijk}^{ABC}) içeren ‘doymuş (saturated) hiyerarşik logaritmik doğrusal model’dir.

Hiyerarşik logaritmik doğrusal modellere göre yüksek dereceli bir terim modelde var ise, daha düşük tüm terimlerin de modelde yer alması gerekmektedir. Yani bu model en karmaşık etkileşim yapısından en basit etkilere kadar ifade edilebilmektedir.

2.3. Çapraz Tablolarda İlişki Ölçütleri

Bu bölümde anlatılan ve çalışmanın analiz kısmında da kullanılan ilişki ölçütleri Agresti (1996)'den yararlanılarak açıklanmıştır.

2.3.1. Göreli Risk

2×2 'lik bir çapraz tabloda göreli risk; iki grup için “başarı” olasılıklarının oranıdır ve negatif olmayan herhangi bir reel sayıya eşittir. Göreli Risk, “GR” olarak ifade edilmekte ve

$$GR = \frac{\pi_1}{\pi_2} \quad (2.1)$$

şeklinde hesaplanmaktadır.

Yığına ait göreli risk ise örnek başarı olasılıklarının oranlanması ile tahmin edilmektedir. İki grup için örnek oranları p_1 ve p_2 olduğunda örnek göreli risk p_1/p_2 olmaktadır.

Örnek göreli riskin doğal logaritması $\log\left(\frac{p_1}{p_2}\right)$, büyük örneklerde beklenen değeri $\log\left(\frac{\pi_1}{\pi_2}\right)$ ve varyansı $\frac{1-\pi_1}{N_1\pi_1} + \frac{1-\pi_2}{N_2\pi_2}$ olan asimptotik normal dağılıma yakınsayacağından,

$\log\left(\frac{p_1}{p_2}\right)$ için $\% (1 - \alpha) \times 100$ Güven Aralığı;

$$\log\left(\frac{p_1}{p_2}\right) \pm z_{\alpha/2} \cdot \sqrt{\frac{1-p_1}{n_1 \cdot p_1} + \frac{1-p_2}{n_2 \cdot p_2}} \quad (2.2)$$

ile hesaplanmaktadır. Bu güven aralığında varyans bilinmediğinden tahmin edicisi kullanılmaktadır. $\log\left(\frac{\pi_1}{\pi_2}\right)$ parametresinin güven aralığının alt sınırı AS ve üst sınırı ÜS olarak tanımlanırsa, $\frac{\pi_1}{\pi_2}$ için $\% (1 - \alpha) \times 100$ güven aralığı, $(e^{AS}, e^{ÜS})$ ile elde edilir (Dağalp, 2018, Tang, He, & Tu, 2012).

2.3.2. Odds Oranı

Odds oranı, kategorik verilerin analizinde uygun model tespit edildikten sonra modeldeki parametrelerin yorumlanmasında sıklıkla kullanılan bir ölçüttür. 2×2 'lik bir çapraz tabloda 1. satır için başarının olasılığı π_1 ve 2. satır için başarının olasılığı π_2 olsun. 1. satır ve 2. satır için başarının odds değerleri sırasıyla $odds_1$ ve $odds_2$ olmak üzere,

$$odds_1 = \frac{\pi_1}{(1-\pi_1)} \quad (2.3)$$

$$odds_2 = \frac{\pi_2}{(1-\pi_2)} \quad (2.4)$$

şeklinde tanımlanmaktadır. Örneğin; $\pi = 0.75$ ise odds değeri, $odds = 0.75/0.25 = 3$ olur. Bu değer başarının gerçekleşme olasılığının başarısızlığa göre 3 kat daha fazla olduğunu göstermektedir. Başarılı olma olasılığı $\pi = 0.8$ ise $odds = 0.8/0.2 = 4.0$ olur ve başarının gerçekleşme olasılığının başarısızlığa göre 4 kat fazla olduğunu, her bir başarısızlık için 4 başarının gözlendiğini ifade etmektedir. Ayrıca herhangi bir olayın meydana gelmesinin odds'u biliniyorsa olayın gerçekleşme olasılığı;

$$\pi = \frac{odds}{odds+1} \quad (2.5)$$

eşitliği ile elde edilir. İki ayrı odds değerinin ya da koşullu odds değerinin birbirine oranı "Çapraz Çarpım Oranı (Cross Product Ratio)" veya "Odds Oranı" olarak tanımlanmakta olup " θ " veya "OR" ile ifade edilmekte ve

$$\theta = \frac{odds_1}{odds_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} \quad (2.6)$$

şeklinde hesaplanmaktadır. Eşitlik (2.1) ve (2.6)'dan anlaşılacağı üzere görel risk iki olasılığın birbirine oranı iken odds oranı θ , iki odds'un birbirine oranını ifade etmektedir.

Odds oranı negatif olmayan herhangi bir reel sayıya eşit olup satır ve sütun değerlerinin yer değiştirmesiyle değişmez, yani simetriktir (invariant). Odds oranının değeri A ve B değişkenleri bağımsız olduklarında, başarı oranları eşit olacağından odds değerleri de eşit olacaktır. Bu durumda odds oranı $\theta = 1$ olacaktır. Bağımsızlığa karşılık gelen 1 değeri, karşılaştırma için bir temel teşkil etmektedir.

Odds oranının en çok olabilirlik tahmin edicisi $\hat{\theta}$,

$$\hat{\theta} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad (2.7)$$

şeklinde ifade edilmektedir.

Küçük ve orta örneklem büyüklükleri için örnek odds oranı oldukça çarpık bir dağılıma sahiptir. Örnek odds oranının logaritmik bir dönüşümü olan $\log \hat{\theta}$, büyük örneklerde ortalaması $\log \theta$ ve asimptotik standart hata (ASE) olarak adlandırılan bir standart sapması ile normal dağılıma yakınsar. 2×2 lik bir çapraz tablo için standart hata,

$$ASE(\log \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \quad (2.8)$$

ile verilmektedir. Bu örnekleme dağılımı normale yakın olduğundan önce $\log \theta$ için güven aralığı hesaplanır ve anti-logaritması alınarak θ için güven aralığı elde edilir.

$\log(\theta)$ için $\%(1 - \alpha) \times 100$ Güven Aralığı;

$$\log(\hat{\theta}) \pm z_{\alpha/2} ASE(\log \hat{\theta}) \quad (2.9)$$

ile elde edileceğinden θ için $\%(1 - \alpha) \times 100$ güven aralığı,

$$AS = \log(\hat{\theta}) - z_{\alpha/2} ASE(\log \hat{\theta}) \quad (2.10)$$

$$\text{ÜS} = \log(\hat{\theta}) + z_{\alpha/2} ASE(\log \hat{\theta}) \quad (2.11)$$

olmak üzere, $(e^{AS}, e^{\text{ÜS}})$ ile hesaplanmaktadır.

Güven aralığı, değişkenlerin birbirinden bağımsız olduğunu ifade eden, $\theta = 1$ değerini içermiyorsa, hesaplanan odds oranının anlamlı olduğu yani odds oranının iki grup arasında farklı olduğu yorumu yapılabilir.

2.3.3. Odds Oranı ve Görel Risk Arasındaki İlişki

Odds oranı ve görel risk her iki grup için başarı oranının sifıra yakın olması durumunda benzer değerler almaktadır. Odds oranı ve görel risk arasındaki ilişki;

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = GR \times \left(\frac{1-p_2}{1-p_1}\right) \quad (2.12)$$

şeklinde tanımlanmaktadır. İkisi arasında tanımlanan bu ilişki bazı durumlarda kullanışlıdır. Örneğin bazı veri setleri için görel riskin hesaplanması mümkün değildir ancak odds oranı hesaplanabilir ve bunun sonucunda görel risk yaklaşık olarak bulunabilir. Bu çalışmada uygun

model tespit edildikten sonra değişkenler arasındaki etkileşim yapısı odds oranı hesaplanarak değerlendirilmiştir.

3. ARAŞTIRMANIN KAPSAMI VE VERİLER

Çalışmada, yazarlarından izin alınarak, “Lise ve Üniversite Öğrencilerinde İntihar Riskini Belirlemeye Yönelik Bir Modelin Sınanması” (Hisli Şahin & Durak Batıgün, 2009) çalışmasına ait veriler kullanılmıştır. Çalışmanın örneklemini Ankara ilinde çeşitli lise ve üniversitelerde okuyan 15-25 yaş arasındaki 2343 öğrenci oluşturmaktadır. Burada katılımcıların demografik özelliklerine ait frekans dağılımları sorulara yanıt verenler üzerinden elde edilmiştir. Bunların 1358’i lise (%58.0), 985’i üniversite (%42.0) öğrencisidir. Örneklemin %56.3 ‘ü kız, %42.5’u erkektir. Katılımcılara ilişkin diğer sosyo-demografik özellikler Tablo 3.1.’de verilmiştir.

Tablo 3.1. Katılımcıların Demografik Özellikleri

		Değişken ve Düzeyleri	N	%
Cinsiyet		Kadın	1320	%56.3
		Erkek	995	%42.5
Yaş		15 – 17	1229	%52.5
		18 – 25	1047	%44.7
Öğrenim Düzeyi		Lise	1358	%58.0
		Üniversite	985	%42.0
Anne Eğitim Düzeyi	Eğitim	Okur-yazar olmayan	63	%2.7
		Okur-yazar olan	64	%2.7
		İlkokul	773	%33.0
		Ortaokul	316	%13.5
		Lise	684	%29.1
		Yüksekokul/Üniversite	421	%19.0
Sosyoekonomik Düzey		Düşük	900	%38.4
		Orta	1000	%42.6
		Yüksek	421	%19.0
Annenin sağ olup olmama durumu		Sağ	2318	%98.9
		Hayatta değil	18	%0.8
Babanın sağ olup olmama durumu		Sağ	2250	%96.0
		Hayatta değil	85	%3.6
Aile Birliği		Birlikte	2106	%89.9
		Ayrı	134	%5.7

İntihar Olasılığı Ölçeği (İÖÖ), 1-4 arasında puanlanan Likert tipi bir ölçek olup Cull ve Gill (1988) tarafından geliştirilmiştir. 36 maddeden oluşan, bireylerin kendilerini değerlendirmelerine yönelik bir ölçektir. İntihar Olasılığı Ölçeği'nden alınan yüksek puanlar bireylerde intihar olasılığının yüksekliğine işaret etmektedir. Formun Türkçeye çevirisi Eskin (1993) tarafından yapılmıştır. Ölçeğin bu çalışmada kullanılan formu ise, Şahin ve Batıgün (2000)'in yapmış olduğu bir çalışmada kullanılan formdur.

Örneklemin İntihar Olasılığı Ölçeği toplam puanı ortalaması, $\bar{X} = 73.07$ ve standart sapması, $ss = 12.25$ olup; intihar olasılığı puanı ortalamasının bir standart sapma üzerinde (85) olanlar “yüksek risk grubu, bir standart sapma altında (61) olanlar ise “düşük risk grubu” şeklinde Hisli Şahin & Durak Batıgün (2009)'de verildiği gibi tanımlanmıştır.

Çalışmada anne eğitim düzeyi, sosyoekonomik düzey göstergesi olarak ele alınmıştır. Annesi okur-yazar olmayan, okur-yazar olan ve ilkokul mezunu olanlar “düşük” (%38.4), ortaokul ve lise mezunu olanlar “orta” (%42.6), yüksekokul veya üniversite mezunu olanlar “yüksek” (%19.0) olarak sınıflandırılmıştır. Ancak logaritmik doğrusal model analizinde sosyoekonomik düzeyi orta grupta yer alan bireyler çalışmaya dahil edilmemiştir. Literatüre bakıldığında, anne eğitim düzeyi ile intihar olasılığı arasında negatif yönde bir ilişki bulunduğuna yönelik bilgiler yer almaktadır (Güleç ve Aksaray, 2006). Anne eğitim düzeyinin, sosyoekonomik düzey göstergesi olarak ele alındığı bu çalışmada da, ters yönlü bir ilişki bulunmuştur. Dolayısıyla anne eğitim düzeyinin intihar olasılığını açıklayan bir değişken olduğu görülmüştür.

4. BULGULAR

Cinsiyet, intihar olasılığı ölçeği (ioö), sosyoekonomik düzey (sed) olmak üzere üç değişken kullanılarak oluşturulan logaritmik doğrusal analiz sonucu Tablo 4.1'de verilmiştir.

Tablo 4.1. Üç Yönlü Etki ve Etkileşim Özet Tablosu

	K	sd	Olabilirlik Oran		Pearson		İterasyon Sayısı
			Ki-Kare	p	Ki-Kare	p	
K-Yönlü ve Daha Yüksek Etkileşimler	1	11	65.638	0.000	68.723	0.000	0
	2	7	21.549	0.003	21.454	0.003	2
	3	2	0.055	0.973	0.055	0.973	3
K-Yönlü Etkileşimler	1	4	44.089	0.000	47.269	0.000	0
	2	5	21.495	0.001	21.400	0.001	0
	3	2	0.055	0.973	0.055	0.973	0

Tablo 4.1'deki ilk kısmın birinci satırı incelendiğinde ana etkiler (K=1), ikili etkileşim etkileri (K=2) ve üçlü etkileşim etkilerinin (K=3) sıfır olduğunu ifade eden H_0 yokluk hipotezi, gerek olabirlik oranı ($L^2 = 65.638, p = 0.000 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 68.723, p = 0.000 < 0.05$) değerlerine göre 0.05 anlam düzeyinde red edilmektedir.

İkinci satırda ise, ikinci dereceden ve üçüncü dereceden etkilerin sıfıra eşit olduğu H_0 yokluk hipotezi, gerek olabirlik oranı ($L^2 = 21.549, p = 0.003 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 21.454, p = 0.003 < 0.05$) değerlerine göre 0.05 anlam düzeyinde yokluk hipotezi red edilmektedir.

Son satırda ise, üçüncü dereceden etkileşim etkisinin sıfıra eşit olduğunu ifade eden H_0 yokluk hipotezi, hem olabirlik oranı ($L^2 = 0.055, p = 0.973 > 0.05$), hem de Pearson Ki-Kare test istatistiği ($\chi^2 = 0.055, p = 0.973 > 0.05$) değerlerine göre 0.05 anlam düzeyinde yokluk hipotezi red edilememiştir. Yani üçüncü dereceden etkileşim etkisi istatistiksel olarak anlamlı değildir.

Tablo 4.1'in ikinci kısmı ise, sırasıyla ana etkiler, ikinci dereceden ve üçüncü dereceden etkilerin sıfıra eşit olduğu hipotezlerini test etmektedir. Ana etkiler ve ikinci dereceden etkileşim etkilerinin sıfıra eşit olduğu H_0 yokluk hipotezi, gerek olabirlik oranı, gerekse Pearson Ki-Kare test istatistikleri p değerleri 0.05 anlam düzeyinde red edilmektedir ($p = 0.000$ ve $p = 0.001 < 0.05$). Ancak üçüncü dereceden etkileşim etkisinin sıfıra eşit olduğu H_0 yokluk hipotezi, gerek olabirlik oranı, gerekse Pearson Ki-Kare test istatistiği p değeri, 0.05 anlam düzeyinde ($p = 0.973 > 0.05$) red edilememektedir. Bu sonuçlar verilere en uygun modelin ana etkiler ve ikinci dereceden etkileşim etkileri parametrelerini içeren ancak üçüncü dereceden etkileşim etkileri parametrelerini içermeyen doymamış hiyerarşik logaritmik doğrusal model olacağını göstermektedir.

Modelde yer alacak ikinci dereceden etkileşim parametrelerini tespit etmek için kısmi ilişkiler test istatistiği sonucu Tablo 4.2.'de verilmiştir.

Tablo 4.2. Etki ve Etkileşimlere Ait Kısmi Ki-Kare ve Olasılık Değerleri

Etki	sd	Kısmi Ki-Kare	p	İterasyon Sayısı
cinsiyet*sed	2	1.422	0.491	2
cinsiyet*ioö	1	22.102	0.000*	2
sed*ioö	2	7.346	0.025*	2
cinsiyet	1	19.179	0.000*	2
sed	2	80.696	0.000*	2
ioö	1	3.029	0.082**	2

*: $p < 0.05$

**.: $p < 0.10$

Tablo 4.2. incelendiğinde ana etkilerden cinsiyet, sosyoekonomik düzey; ikinci dereceden etkileşim etkisi parametrelerinden cinsiyet*ioö, sed*ioö etkileşim etkileri 0.05 anlam düzeyinde istatistiksel olarak anlamlı bulunmuştur. Ancak hiyerarşik logaritmik doğrusal modellere göre yüksek dereceli bir terim modelde var ise, daha düşük tüm terimlerin de modelde var olması gerekmektedir. Bu nedenle kısmi ilişkiler tablosunda 0.05 anlam düzeyinde anlamlı bulunmayan (ioö) ana etki parametresinin, cinsiyet ve sosyoekonomik düzey değişkenleri arasında ikili etkileşim etkisinin olması nedeni ile modele dâhil edilmesi gerektiği tespit edilmiştir. Bu durumda A, B ve C kategorik değişkenleri için verilere en uygun model;

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} , \quad i = 1,2 , j = 1,2 \text{ ve } k = 1,2$$

şeklinde ifade edilen koşullu bağımsızlık modeli olarak bulunmuştur.

Ana etkiler ve etkileşimlere ait parametre tahminleri aşağıdaki tabloda özetlenmiştir.

Tablo 4.3. Etki ve Etkileşimlere Ait Parametre Tahminleri

Etki	Parametre	Tahmin	Std. Hata	Z	p	%95 Güven Aralığı	
						Alt Sınır	Üst Sınır
cinsiyet*ioö*sed	1	0.022	0.092	0.236	0.814	-0.159	0.203
	2	-0.012	0.095	-0.129	0.897	-0.199	0.174
cinsiyet*ioö	1	0.207	0.071	2.933	0.003*	0.069	0.345
cinsiyet*sed	1	-0.125	0.092	-1.353	0.176	-0.306	0.056
	2	0.269	0.095	2.822	0.005*	0.082	0.455
ioö*sed	1	-0.106	0.092	-1.143	0.253	-0.287	0.075
	2	-0.076	0.095	-0.802	0.423	-0.263	0.110
Cinsiyet	1	0.235	0.071	3.338	0.001*	0.097	0.374
İoö	1	0.061	0.071	0.864	0.388	-0.077	0.199
Sed	1	0.243	0.092	2.631	0.009*	0.062	0.424
	2	0.212	0.095	2.231	0.026*	0.026	0.399

*: p < 0.05

Tablo 4.3 incelendiğinde uygun modelde yer alan bazı değişkenlerin (cinsiyet, sed, cinsiyet*ioö) p değerlerinin 0.05'den küçük olduğu görülmektedir.

Ayrıca standartlaştırılmış parametre tahminleri göz önünde bulundurularak Tablo 4.3 incelendiğinde ana etkiler arasında en büyük Z değeri 3.338 olup, cinsiyet parametresinin kadın kategorisine ait olduğu görülmektedir. Yani değişkenler arasında hücre frekansına en önemli katkıyı sağlayan faktör cinsiyet değişkenidir. İstatistiksel olarak anlamlı bulunan ikili etkileşim etkilerinin standartlaştırılmış parametre tahmin değerlerine bakıldığında ise intihar olasılığı düşük riskli grupta yer alanların yine cinsiyet değişkeninin kadın kategorisinde yer alanlara bağımlı olduğu; sosyoekonomik düzeyi yüksek olanların cinsiyet değişkeninin yine kadın kategorisinde bağımlı olduğu görülmektedir.

Cinsiyet ve intihar olasılığı ölçeği etkileşim parametreleri dikkate alındığında cinsiyeti kadın olanların intihar olasılığı düşük olan grupta olma oranı, erkeklerin intihar olasılığı düşük grupta olma oranından 2.13 kat daha fazladır. Örnek odds oranı kullanılarak hesaplanan %95 güven aralığı (1.56, 2.92), değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içermediğinden hesaplanan odds oranının istatistiksel olarak anlamlı olduğu dolayısıyla gruplar arasında farklılık olduğu tespit edilmiştir. Ayrıca sosyoekonomik düzeyi yüksek olan grubun intihar olasılığı düşük grupta olma oranı, sosyoekonomik düzeyi düşük olan grubun intihar olasılığı düşük grupta olma oranından 1.63 kat daha fazla olduğu bulunmuştur. Ancak örnek odds oranı kullanılarak hesaplanan %95 güven aralığı (0.40,1.06), değişkenlerin birbirinden bağımsız olduğunu ifaden eden $\theta = 1$ değerini içermediğinden hesaplanan odds oranının istatistiksel olarak anlamlı olmadığı yani gruplar arasında anlamlı farklılık olmadığı tespit edilmiştir.

5. SONUÇ

Bu çalışmanın amacı, kategorik verilerin analizinde kullanılan ve bağımlı-bağımsız ayrımı yapmadan değişkenler arasındaki ana etkileri, ikili ve daha yüksek dereceden etkileşim etkilerini ortaya çıkaran, aynı anda birden fazla hipotezin test edilmesine olanak sağlayan logaritmik doğrusal modeller ile cinsiyet, sosyoekonomik düzey ve intihar olasılığı değişkenleri arasındaki etkileşim yapısının incelenmesidir. Bu amaçla, Hisli Şahin ve Durak Batıgün (2009) tarafından gerçekleştirilen “Lise ve Üniversite Öğrencilerinde İntihar Riskini Belirlemeye Yönelik Bir Modelin Sınanması” çalışmasına ait 15-25 yaş gurubu arasındaki 2343 öğrenciden elde edilen veriler kullanılmıştır.

Cinsiyet, sosyoekonomik düzey ve intihar olasılığı değişkenleri olmak üzere üç değişken kullanılarak oluşturulan logaritmik doğrusal analiz sonucuna göre ana etkiler ve ikinci dereceden etkileşim etkilerinin sifira eşit olduğu hipotezi red edilmiş ancak üçüncü dereceden etkileşim etkilerinin sifira eşit olduğu hipotezi red edilememiştir. Buna göre nihai modelin ikinci dereceden etkileşim etkilerini içeren ancak üçüncü dereceden etkileşim etkilerini içermeyen doymamış logaritmik doğrusal model olduğuna karar verilmiştir.

Standartlaştırılmış parametre tahminleri incelendiğinde ilgili değişkenler arasında hücre frekansına en önemli katkıyı sağlayan faktör cinsiyet değişkenidir. Ayrıca intihar olasılığı düşük grupta yer alanların cinsiyet değişkeninin kadın kategorisinde yer alanlara bağımlı olduğu; sosyoekonomik düzeyi yüksek olanların da cinsiyet değişkeninin yine kadın kategorisinde bağımlı olduğu tespit edilmiştir. İkili etkileşim etkileri incelendiğinde ise kadınların intihar olasılığı düşük grupta olma oranı, erkeklere göre 2.13 kat daha fazla olarak bulunmuştur. Bu sonuç erkeklerin kadınlara göre intihar olasılığı yüksek olan grupta olma oranının daha fazla olduğunu göstermektedir. Ayrıca istatistiksel olarak da gruplar arasında anlamlı farklılık olduğu tespit edilmiştir. Anne eğitim düzeyinin sosyoekonomik düzey göstergesi olarak ele alındığı bu çalışmada, sosyoekonomik düzeyi yüksek olanların intihar olasılığı düşük grupta olma oranı, sosyoekonomik düzeyi düşük olanların intihar olasılığı düşük grupta olma oranından 1.63 kat daha fazla bulunmuştur. Bu sonuç, literatürde anne eğitim düzeyi ile intihar olasılığı arasında

negatif yönde bir ilişki bulunduğuna yönelik bilgileri destekler niteliktedir. Ancak istatistiksel olarak gruplar arasında anlamlı farklılık olmadığı tespit edilmiştir.

Bu bulgular doğrultusunda öncelikle lise ve üniversite öğrencilerinin intihar düşünce ve davranışları açısından değerlendirilmesinin önemli olduğu söylenebilir. Bu yaş grubundaki öğrencilere yönelik psikolojik destek sağlayan kurum ve kuruluşların sayılarının artırılması ve son olarak ülkemizin intihar olasılığını çok yönlü ele alarak politika süreçlerine bu anlamda yön vermesi önerilebilir. Ancak çalışmanın önemli bir kısıtlılığı, elde edilen bulguların, sadece Ankara ilinde çeşitli lise ve üniversitelerde okuyan öğrenci kitlesine ait olup, tüm lise ve üniversite öğrencilerine genellenemeyecek olmasıdır. Ayrıca logaritmik doğrusal modellerde değişkenler arasında açıklanan ve açıklayıcı değişken ayrımı yapılmadığından, elde edilen bulgular ilişki düzeyinde olup, değişkenler arasında neden-sonuç ilişkisi vermemektedir. Bu nedenle çalışmanın daha geniş bir örneklem grubuyla ve bireylerin intihar olasılığında risk faktörü olduğu düşünüldüğü başka değişkenlerinde analiz çalışmasına dahil edilerek incelenmesi sonucunda fayda sağlayabileceği düşünülmektedir.

KAYNAKÇA

- Agresti, A. (1990). *Categorical Data Analysis*, John Wiley & Sons Inc, New York.
- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley and Sons, Inc., Canada.
- Andersen, E. (1990). *The statistical analysis of categorical data*. Springer-Verlag, Berlin, 520 (1990).
- Atay, İ. M., Eren, İ., ve Gündoğar, D. (2012) *Isparta İl Merkezinde İntihar Girişimi, Ölüm Düşünceleri Yaygınlığı ve Risk Faktörleri*.
- Bülbül, S. (2006). *Üç Boyutlu Çapraz Tablolarda Logaritmik Doğrusal Analiz: Çocuk İşgücü Değişkenleri Arasındaki Etkileşimler*. Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , 41-70.
- Cull, JG, & Gill, WS (1988). *Suicide Probability Scale (SPS) Manual*. . Western Psychological Services, Los Angeles.
- Dağalp, R. (2018). *ST 431 (Klinik Deneylerde İstatistiksel Yöntemler) Ders Notları*. Ankara Üniversitesi, Açık Ders Malzemeleri: <https://acikders.ankara.edu.tr/course/view.php?id=5015>, Erişim Tarihi: 15.11.2018.
- Erdem, A. (2014). *Uygunluk Analizinde Logaritmik Doğrusal Modellerin Kullanımı: Televizyon İzleme Eğilimleri Üzerine Bir Uygulama*. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Yüksek Lisans Tezi, Haziran 2014.
- Erdugan, F., ve Türkan, A. H. (2017). *Üç Yönlü Kontenjans Tablolarında Log-Lineer Model ile İş Kazası Verilerinin İncelenmesi*. Karaelmas Fen ve Mühendislik Dergisi, 462-468.
- Eskin, M. (1993). *Age specific suicide rates and the rates of increase, and suicide methods in Sweden and Turkey. A comparison of the official suicide statistics*. Reports from the Department of Psychology, Stockholm University , No: 772.
- Filiz, Z. (2007). *Üç Yönlü Log-Lineer Modeller İle Üniversite Öğrencilerinin Sigara, Alkol ve Nargile İçme Nedenlerini Etkileyen Faktörlerin Belirlenmesi*. DergiPark Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 8(2) , 225-250.
- Goodman, L. A. (1970). *"The multivariate analysis of qualitative data:interaction among multiple classification"*. J.Amer.Statist.Assos.,65,226-256 (1970).
- Güleç, G., ve Aksaray, G. (2006). *İntihar Girişiminde Bulunan Gençlerin Sosyodemografik-Sosyokültürel ve Aile Özelliklerinin Değerlendirilmesi*. In Yeni Symposium, Türk Psikiyatri Dizini (Vol. 44, No. 3, pp. 141-150).
- Haberman, S. J. (1974). *"Loglinear models for frequency tables with ordered classifications"*. Biometrics, 36:589-600 (1974).
- Hisli Şahin, N., ve Durak Batıgün, A. (2009). *Lise ve Üniversite Öğrencilerinde İntihar Riskini Belirlemeye Yönelik Bir Modelin Sınanması*. Türk Psikiyatri Dergisi 2009; 20(1):28-36.
- Jeansonne, A. (2017). *Loglinear Models*. <http://userwww.sfsu.edu/efc/classes/biol710/loglinear/Log%20Linear%20Models.pdf>, Erişim Tarihi: 02.01.2017.

- Kaşkır, F. (2012). *Logaritmik Doğrusal Modeller ve Uygunluk Analizinin Birlikte Kullanımı: Lise Öğrencilerinin Sigara İçme Alışkanlıklarına Uygulanması*. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Eylül 2012.
- Köleoğlu, N. (2018). *Uluslararası Öğrencilerin Sosyal Uyum Problemini Belirlemede Loglineer Model Kullanımı*. Uluslararası İktisadi ve İdari İncelemeler Dergisi, UIİİD-IJEAS, 2018 (18. EYİ Özel Sayısı):101-116.
- Mete, S., ve Ünsal, A. (2010). *Kategorik Veriler İçin Logaritmik Doğrusal Modeller ve Göç İstatistikleri Üzerine Bir Uygulama*. DergiPark Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 9-20.
- Sayıl, I., Berksun O., Palabıyıkoglu R., Oral A., Haran, S. Güney, S. Binici, S. Geçim S., Yücat T., Beder A., Özayar H., Büyükelik D. ve Devrimci Özgüven H., “*Attempted Suicides in Ankara in 1995*”, Crisis, 19, 1, 47-48 (1998).
- Sayıl, I. ve Devrimci Özgüven H., “*Suicide and Suicide Attempts in Ankara in 1998: Results of The Who/Euro Multicentre Study on Suicidal Behavior*” Crisis, 23, 1, 11-16 (2002).
- Simons, R. L., & Murphy, P. I. (1985). *Sex differences in the causes of adolescent suicide ideation*. *Journal of Youth and Adolescence*, 14, 423 – 434
- Şahin, N., ve Batıgün, AD. (2000). *Yaşamı sürdürme nedenleri ve intihar olasılığı*, (Yayınlanmamış Çalışma).
- Şimşek Kandemir, A., ve Şimşek, M. (2019). *Geçici İş Göremezlik Süresi, Cinsiyet, Çalışılan Ortam ve İş Kazası Türleri Etkileşimlerinin İstatistiksel Analizi*. Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi, 7 (1) 239-245.
- Tang, W., He, H., & Tu, X. M. (2012). *Applied Categorical and Count Data Analysis*. Boca Raton, London, New York: CRC Press, Taylor & Francis Group.
- Uygun, H. (1990). *Çapraz Tabloların Çözümlemesi ve Log-Linear Modeller*. Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:8, Sayı:1 , (s:299-308).
- Yılmaz, V., ve Şıklar, E. (2002). *İntiharlarla İlgili Çok Değişkenli Kategorik Verilerin Analizinde Logaritmik Doğrusal Modellerin Kullanılması*. *Anadolu Üniversitesi, Bilim ve Teknoloji Dergisi* , Cilt.3 , Sayı.2 , 271-280 .
- Yurt Öncel, S., ve Erdugan, F. (2015). *Kontenjans Tablolarının Analizinde Log-Linear Modellerin Kullanımı ve Sigara Bağımlılığı Üzerine Bir Uygulama*. Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi Cilt 19 Sayı 2 , 221-235.

Extended Summary

Use of Logarithmic Linear Models in Analysis of Categorical Data: An Application on Suicide Risk Data

Suicide is defined by the Turkish Language Association (TDK) as the ending of one's own life with the effect of social and spiritual reasons. The World Health Organization (WHO) defines suicide action as "self-harming awareness of the purpose of the person and to varying degrees of mortality". According to a report published by the World Health Organization, one person commits suicide every 40 seconds in the World. Suicide is shown to be the second biggest cause of death after road accidents among young people aged 15-29. In addition, according to released data by Turkey Statistical Institute (TSI) in 2018, 3161 people committed suicide; in other words, it is revealed that an average of 9 people ends their lives every day. Young people between the ages of 15-24 are considered to be a risky group in terms of suicide both in the World and in our country (Atay, Eren ve Gündoğar, 2012; Sayıl ve ark., 1998; Sayıl, Devrimci Özgüven 2002; Simons ve Murphy, 1985). Therefore, it is important to evaluate the probability of suicide in the risk group in order to take precautions against increasing suicide rates. The aim of this study is to determine the probability of suicide for students aged 15-25; to examine the interaction structure with variables such as gender and socioeconomic level which are considered among the risk that may lead the individual to suicide.

The data obtained as a result of researches in social sciences are generally categorical data. However, the historical development process of the methods used in the analysis of categorical data is not very old. The development of statistical methods that will be applied to categorical data until the 1960s is behind the development of methods applied to quantitative data. Effective studies were carried out in this field by Karl Pearson in the 1900s, and studies dealing with Logarithmic linear model applications started to develop after the 1960s, in the analysis of categorical data with the studies of scientists such as Goodman (1970), Haberman (1974), Agresti (1990). The use of Logarithmic Linear (Log-Linear) Model applications has gained importance.

The sample of the study consists of 2343 students between the ages of 15-25 who study in various high schools and universities in Ankara. 1358 of them are high school students (58.0%) and 985 are university students (42.0%). 56.3% of the sample is female and 42.5% is male. In this context, the interaction structure of variables consisting of gender, socioeconomic level and probability of suicide was examined with Logarithmic Linear Models by using data from the study of "Testing a Model for Determining Suicide Risk in High School and University Students" (Hisli Şahin & Durak Batgün, 2009).

While only independence is tested in two-way cross tables in which two categorical variables are addressed, in the analysis of cross tables formed by three or more variables, hypotheses related to basic and interaction effects are also be tested. For this reason, logarithmic linear model analysis has an important place among discrete multivariate analysis and its use is common in applications. All variables examined in logarithmic linear models are treated as response variables. In other words, there is no discriminating variable among the explained variables. Therefore, logarithmic linear models only reveal the relationship structure between variables (Jeansonne, 2017). However, if it is desired to differentiate the response and explanatory variables, a logit or logistic regression model should be used instead of logarithmic linear models.

Based on the results of the three-way logarithmic linear analysis, in which the probability of suicide is investigated for students between the ages of 15-25 who study in various high schools and universities, the hypothesis that third-order interaction effects are equal to zero

could not be rejected but the main effects and the second-order interaction effects are found significant. According to this model, it was determined that the value that provides the most important contribution to the frequency of the cell belongs to the female category, and those who are in the risk group with a low risk of suicide are again dependent on the female category of the gender variable. The rate of women in the low-risk suicide group was 2.13 times higher than in men. In other words, men are more likely to be in a high risk suicide group than women. Since the 95% Confidence Interval calculated using the sample odds ratio (2.135) does not contain the value that indicates the variables are independent of each other, the calculated odds ratio is statistically significant and therefore there is a difference between the groups. In addition, the rate of those with high socio-economic status in the low-risk suicide group was 1.63 times higher than those with low-socio-economic status in the low-risk suicide group. However, since the 95% Confidence Interval calculated using the sample odds ratio includes the value that indicates independency of variables, it was concluded that the calculated odds ratio is not statistically significant, that is, there is no difference between the groups.

In line with these findings and considering the risk of potential young suicide in our country, to protect the high school and university students from risk of suicide, it is highly recommended that the authorities a) have to take the necessary precautions, and b) must increase the number of institutions and organizations which provide psychological support to students. The sample of this study consists of students studying only in various high schools and universities in Ankara and the narrow age group restricts the generalizability of the findings to all high school and university students. For this reason, it is thought that the study may be beneficial as a result of the study with a larger sample group and other variables that are thought to be a risk factor in the suicide probability of individuals.