T.C
SÜLEYMAN DEMİREL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI

LOGARİTMİK DOĞRUSAL MODELLER VE BİR UYGULAMASI

Esra TOPALOĞLU
1630227036

YÜKSEK LİSANS TEZİ

DANIŞMAN
Dr. Öğr. Üyesi Aliye ATAY KAYIŞ

ISPARTA – 2018

TEZ SAVUNMA SINAV TUTANAĞI

T.C.

SÜLEYMAN DEMİREL ÜNİVERSİTESİ

Sosyal Bilimler Enstitüsü Müdürlüğü

YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum "Logaritmik Doğrusal Modeller ve Bir Uygulaması" adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadar ki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya'da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim.

Esra TOPALOĞLU
18/12/2018

(TOPALOĞLU, Esra, *Logaritmik Doğrusal Modeller ve Bir Uygulaması*, Yüksek Lisans Tezi, Isparta, 2018)

ÖZET

Kategorik verilerde istatistiksel yöntemlerin kullanımı sosyal bilimlere ilişkin uygulamalarda oldukça yaygındır. Kategorik değişkenler arasındaki ikili ilişkilerin test edilmesinde Pearson χ^2 test istatistiği yeterli olmakta ancak ikiden fazla değişken olması durumunda bu istatistik kullanılamamaktadır. Bu nedenle üç veya daha çok boyutlu çapraz tablolarda ilişki yapılarının belirlenmesinde χ^2 istatistiğinden farklı olarak değişkenler arasında bağımlı ve bağımsız ayrımı yapmayan "Logaritmik Doğrusal Modeller" kullanılmaktadır. Logaritmik Doğrusal Modeller yardımıyla daha çok değişkenin etkileri analiz edilmekte ve sadece değişkenlere ait ana etkiler değil değişkenler arasındaki ikili veya daha yüksek dereceden etkileşim etkileri de sorgulanabilmektedir. Bu çalışmada Ankara ilinde çeşitli lise ve üniversitelerde okuyan öğrencilere ilişkin değişkenler (demografik, sosyo-demografik, intihar riski vb.) arasındaki etkileşimler "Logaritmik Doğrusal Modeller" yardımı ile incelenmeye çalışılmıştır.

Anahtar Kelimeler : Kategorik Veri, Çapraz Tablolar, Logaritmik Doğrusal Modeller

(TOPALOĞLU, Esra, *Log-Linear Models And An Application*, Yüksek Lisans Tezi, Isparta, 2018)

ABSTRACT

The use of statistical methods in categorical data is quite common in social science applications. The χ^2 statistic of Pearson is sufficient to test the binary relationships between categorical variables. However, this statistical analysis cannot be used if there are more than two variables. Therefore, Logarithmic Linear (Log-Linear) Models which do not discriminate between dependent and independent variables on the contrary to the χ^2statistic are used in the determination of the relationship structures in three or more dimensional cross tables. In other words, by Log-Linear Models not only the main effects of the variables but also two way or higher degree interaction effects among the variables can be investigated. In this study, the interaction effects between variables (such as; demographic, socio-demographic, suicidal risk, etc.) of students in various high schools and universities in Ankara were investigated using Logarithmic Linear Models.

Keywords: Categorical Data, Cross-Tables, Log-Linear Models

İÇİNDEKİLER

ÜÇÜNCÜ BÖLÜM
İNTİHAR OLASILIĞI ÖLÇEĞİ VE SOSYO-DEMOGROFİK DEĞİŞKENLERİN LOGARİTMİK DOĞRUSAL MODELLER YARDIMI İLE İNCELENMESİ

KISALTMALAR

AIC	: Akaike Bilgi Kriteri (Akaike's Information Criterion)
ANOVA	: Varyans Analizi (Analysis of Variance)
ASE	: Asimptotik Standart Hata
AS	: Alt Sınır
GR	: Göreli Risk
İOO	: İntihar Olasılığı Ölçeği
N	: Yığın Çapı
OR	: Odds Oranı
PÇE	: Problem Çözme Envanteri
s	: Örneklem Genişliği
sd	: Serbestlik Derecesi
ss	: Standart Sapma
Std. Hata	: Standart Hata
SPSS	: Statistical Package for the Social Sciences
ÜS	: Üst Sınır
vb.	: Ve benzeri

TABLOLAR

ŞEKİLLER

ÖNSÖZ

"Logaritmik Doğrusal Modeller ve Bir Uygulaması" adlı tez çalışmamın planlanmasında, araştırılmasında ve yürütülmesinde desteklerini benden esirgemeyen, çalışmamla yakından ilgilenen ve eleştirileriyle yol gösteren kendisiyle çalışmaktan onur duyduğum sevgili danışman hocam Dr. Ögr. Üyesi Aliye ATAY KAYIŞ'a en derin teşekkürlerimi sunarım.

Ayrıca, çalışmamın analizinde kullanmış olduğum veriyi temin eden Ankara Üniversitesi, Psikoloji Bölümü, Öğretim Üyesi Prof. Dr. Ayşegül DURAK BATIGÜN'e, çalışmama bilgileriyle katkı sağlayan Ankara Üniversitesi, İstatistik Bölümü, Öğretim Üyesi Doç. Dr. Rukiye DAĞALP'e ve çalışmamda motivasyonumu yüksek tutmama yardımcı olan Süleyman Demirel Üniversitesi Maliye Bölümü Dr. Öğr. Üyesi Ceyda ŞATAF'a teşekkürlerimi sunarım.

Son olarak, benimle birlikte aynı heyecanı ve stresi paylaşan, bana her zaman maddi ve manevi destek olan, başta annem, babam ve kardeşlerim olmak üzere bu süreçte benimle yakından ilgilenen tüm dostlarıma en içten teşekkürlerimi sunarım.

GİRİŞ

Bir araştırmanın istatistiksel değerlendirme süreci yapılırken verilere uygun istatistiksel teknikler kullanılması önemlidir. Ayrıca, ilgilenilen veriler için hangi ölçeğin kullanılacağı en az uygun istatistiksel tekniklerin belirlenmesi kadar önemlidir. İstatistiksel analiz çalışmalarında ele alınan değişkenler arasındaki ölçüm düzeyleri nominal (isimsel), ordinal (sıralı), interval (aralıklı) veya ratio (oransal) şeklinde olabilir. Bu ölçüm düzeylerine göre de kullanılacak istatistiksel teknikler farklılık gösterecektir. Örneğin; ölçüm düzeyi nominal veya ordinal olan verilerde parametrik olmayan (nonparametrik) istatistiksel analiz yöntemleri kullanılırken, ölçüm düzeyi aralıklı veya oransal olan verilerde parametrik istatistiksel analiz yöntemleri kullanılmaktadır. Parametrik olmayan istatistiksel yöntemler parametrik istatistik yöntemlere göre daha esnek istatistiksel yöntemlerdir.

Parametrik yöntemlerin kullanılabilmesi için gözlem sayısının yeterince büyük olması, verilerin normal dağılması ve varyansların homojenliliği gibi bazı varsayımların sağlanması gerekirken, parametrik olmayan yöntemlerde ise benzer varsayımlar aranmamakta, örneklem dağılımına ilişkin varsayımlar ortaya konmamaktadır.

Sosyal bilim alanlarında yapılan araştırmalar sonucu elde edilen veriler genellikle kategorik veriler olup, parametrik analiz varsayımları sağlanmadığından parametrik olmayan yöntemler kullanılmaktadır. Bu nedenle bu tür verilerin analizinde kullanılan en yaygın metot genelleştirilmiş logaritmik doğrusal modellerin bir uzantısı olan ‘Logaritmik Doğrusal Modeller’dir.

Logaritmik doğrusal modellerde gözlem sonucu elde edilen veriler çeşitli çapraz-sınıflandırılmış tablolar halinde düzenlenerek analizler yapılır ve yorumlanır. Kategorik verilerle yapılan çalışmalardaki temel amaç değişkenler arasındaki karmaşık ilişki yapılarını ortaya çıkartmaktır. Bu amaç doğrultusunda çapraz tabloların boyutlarını oluşturan değişkenler arasındaki ana etkiler, ikili ve daha üst düzey etkileşim etkileri en uygun logaritmik doğrusal modelle ifade edilir.

Çapraz tablolar yardımı ile kategorik verilerin analizinde kullanılan logaritmik doğrusal modellerin kullanımı, sosyal bilimlerde sıklıkla kullanılmakla birlikte, sağlık bilimlerinde, davranış bilimlerinde, halk sağlığında, pazarlamada ve daha birçok alanda da kullanımı oldukça yaygındır.

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 ve logaritmik doğrusal model uygulamalarını ele alan çalışmalar 1960'lı yıllardan sonra gelişmeye başlamış, Goodman (1970), Haberman (1974), Agresti (1990) gibi bilim adamlarının çalışmaları ile Logaritmik Doğrusal (Log-Linear) Model uygulamaları önem kazanmıştır. Günümüzde de önemini korumaktadır.

Çok değişkenli kategorik verilerin analizinde kullanılan logaritmik doğrusal modellerle ilgili Türkiye'de yapılmış bazı çalışmalar aşağıdaki gibidir.

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. Araştırma sonucu elde edilen en uygun logaritmik doğrusal model ana etkiler ve ikili etkileşimleri içeren doymamış model olarak belirlenmiştir.

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. Bu çalışmada günde içilen sigara sayısının cinsiyet, yaş, barınma şekli ve sigara cinsinden etkilendiği, kardeş sayısı ve sosyal durumdan etkilenmediği bulunmuştur. Alkol kullanım sıklığının, barınma şekli ve alkol türünden etkilendiği, kardeş sayısı, ailenin sosyal durumu, cinsiyet ve yaştan etkilenmediği bulunmuştur. Ayrıca nargile kullanım sıklığının ise cinsiyet, yaş ve barınma şeklinden etkilendiği belirlenmiştir.

Mete ve Ünsal (2010) logaritmik doğrusal modeller ile göç istatistikleri üzerine bir uygulama yapmış ve ele alınan üç-yönlü tablonun, logaritmik doğrusal modelle analiz edilmesi sonucunda ikinci ve üçüncü dereceden etkileşimlerin istatistiksel olarak anlamlı olduğu doymuş bir model elde edilmiştir.

Kaşkır (2012) sigara içen lise öğrencilerinin sigara içmelerinde etkili olan nedenleri ortaya çıkarmak amacıyla ele aldığı değişkenler arasındaki ikili ve üçlü etkileşimleri logaritmik doğrusal modeller ile incelenmiştir. Logaritmik doğrusal model uygunluk analizi yapılarak sonuçlar yorumlanmıştır.

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. Kategorik değişkenlerin analizinde kullanılan diğer bir yöntem olan uygunluk analizini değişkenler arasındaki etkileşim ve ilişkileri ortaya koymak amacıyla birbirlerinin tamamlayıcısı olarak kullanmıştır.

Yurt Öncel ve Erdugan (2015) logaritmik doğrusal modeller ile sigara bağımlılığı üzerine bir uygulama yapmıştır. Aynı zamanda iki ve üç yönlü çapraz tabloların logaritmik doğrusal modeller ile analiz edilmesinin klasik çapraz tablo analizlerine göre üstünlüklerini incelemişlerdir. Çalışma sonucunda, Kırıkkale'de yaşayan ve sigara kullanan bir bireyin, sigara bağımlılık düzeyinin orta, cinsiyetinin erkek ve ebeveynlerinin birinin sigara içmesinin logaritmik doğrusal modeldeki frekansı artıran önemli bir etken olarak bulunmuştur.

Erdugan ve Türkan (2017) iş kazalarını, üç yönlü çapraz tablolar halinde düzenleyerek logaritmik doğrusal analiz yöntemi aracılığı ile incelemiş ve analiz sonucu sektör, cinsiyet ve iş göremezlik süresi değişkenlerinin iş kazası sıklığını açıkladığını saptamıştır. Cinsiyet değişkeni için erkek çalışanlar, sektör değişkeni için inşaat sektörü, iş göremezlik değişkeni için ise bir gün kategorisi iş kazası sıklıklarını belirleyici niteliktedir.

Bu çalışmanın da temel amacı kategorik verilerin analizinde kullanılan, değişkenler arasındaki ikinci veya daha yüksek dereceden etkileşim etkilerini logaritmik doğrusal modeller ile test etmektir. Bu kapsamda;

Tezin ilk bölümünde kategorik verilerin analizinde kullanılan çapraz tablolar hakkında temel kavramlar verilmeye çalışılmış, iki ve üç yönlü çapraz tablolar kısaca tanıtılmıştır.

Tezin ikinci bölümde logaritmik doğrusal modeller iki yönlü ve üç yönlü çapraz tablolar için ayrıntılı bir şekilde ele alınmış ve çapraz tablolarda ilişki ölçütlerine yer verilmiştir.

Tezin üçüncü bölümünde ise Logaritmik Doğrusal Analiz yöntemine geçilmeden önce veri grubuna ilişkin özet istatistiklere ve klasik istatistiksel analiz yöntemlerine yer verilmiştir. Daha sonra intihar olasılığı ölçeği, problem çözme envanteri, cinsiyet, sosyo-ekonomik düzey, öğrenim düzeyi gibi çeşitli sosyo-demogrofik değişkenler arasındaki ikili etkileşim etkileri ve daha yüksek dereceden etkileşim etkileri çalışmanın temel amacını oluşturan "Logaritmik Doğrusal Modeller" yöntemi ile analiz edilmiş ve değişkenler arasındaki ilişkiler odds oranları ile yorumlanmıştır.

BİRİNCİ BÖLÜM

KATEGORİK VERİLERİN ANALİZİ VE ÇAPRAZ TABLOLAR

1.1. KATEGORİK DEĞİŞKEN

Sosyal Bilimlerde yapılan araştırmalarda karar vermeye yönelik birçok sayısal araştırma yöntemi mevcuttur. En uygun araştırma yöntemini bulmada en önemli etkenlerden biri de araştırmada kullanılan değişkenlerin yapısıdır.

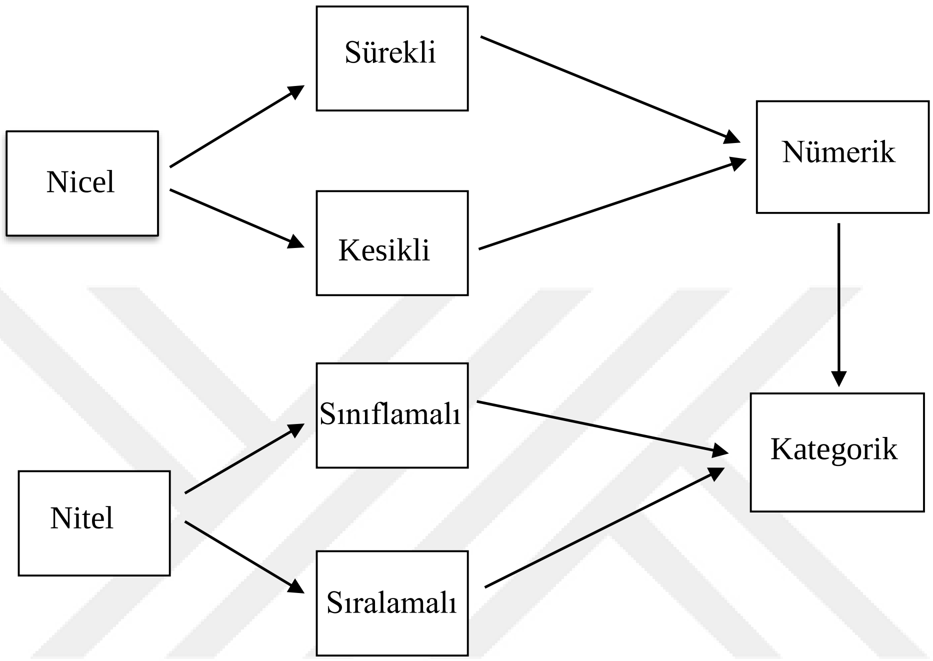
Değişken kavramı genel itibari ile birimlerin ölçülebilen özellikleri olarak tanımlanabilir. Doğrudan sayılarla ifade edilebilme durumlarına göre değişkenler nicel (kantitatif) ve nitel (kalitatif) olmak üzere ikiye ayrılmaktadır. Nitel değişkenler genellikle sosyal bilimlerde karşımıza çıkmakta olup sözcüklerle ifade edilirken, nicel değişkenler sayılarla ifade edilmektedir. Bir anket formundaki kişilerin adı, cinsiyeti, mesleği nitel değişkenlere, kilo ve boy uzunlukları nicel değişkenlere örnek olarak verilebilir.

Nicel değişkenler de kendi aralarında kesikli ve sürekli olmak üzere iki alt gruba ayrılır. Sürekli nicel değişken belirli bir aralıkta tüm noktaları reel bir değer alabiliyorken, kesikli nicel değişken sadece belirli noktalarda değer alabilir. Literatürde kesikli nicel değişkenler sürekli olmayan ancak sayılabilir veriler anlamına gelen “Sayım Verileri (Count Data)” olarak adlandırılır. Nitel değişkenler ise sıralama ve sınıflama ölçme düzeylerine göre ifade edilir (Mete, 2009).

Kategorik değişken; sınıflandırılmış bir veri setini içeren bir ölçü derecesidir. Eğitim durumu, öğrenim düzeyi, cinsiyet vb. değişkenler örnek olarak verilebilir. Ancak sürekli değişkenler kategorik değişkene dönüştürülebilir. Sürekli bir değişkenin kategorik veriye dönüştürülmesinde değişken değerlerinin tekrar eden değerler olmaları önem taşımaktadır (Powers & Xie, 2000).

1.2. KATEGORİK VERİ

Sınıflama (nominal) ve sıralama (ordinal) ölçüm düzeyine sahip veri "Kategorik Veri" olarak tanımlanır. Ölçüm düzeylerine göre verinin sınıflandırılması Şekil 1.1'de verilmiştir.



Şekil 1.1. Verinin Sınıflandırılması

Şekil 1.1'de görüldüğü gibi sayısal olarak ifade edilen sürekli bir değişken, gerektiğinde sınıflayıcı ve sıralayıcı ölçüm düzeyine indirgenerek kategorik hale getirilebilir.

Kategorilerine ayrılmış değişkenlere göre toplanan verilerde, her bir hücreye (gözeye) kaç tane gözlem düştüğü sayım yoluyla belirlenir. Bu nedenle kategorik veri, frekans veri olarak da isimlendirilir (Özdil 2009, Howell 2010, Öztürk 2011, Çilan 2013).

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 yapılı ilişkilerin incelenmesinde "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.

1.3. ÇAPRAZ TABLOLAR

Kategorik değişkenler arasındaki ilişkilerin ortaya çıkarılmasında çapraz tablolar diğer bir ifade ile kontenjans tabloları ya da olumsallık tabloları kullanılmaktadır (Agresti, 1990).

Kategorik değişkenler, satır ve sütunlarda değişken düzeylerinin, hücrelerde ise gözlenen sıklıkların yer aldığı, değişken sayısı k olmak üzere göre iki yönlü, üç yönlü veya k yönlü çapraz tablolar halinde gösterilebilir. Bu tablolardaki hücrelerde yer alan değerler frekans verileri yani sayım verileridir, değişken kategorilerinin veri kümesinde kaç kez tekrarlandığını gösterir.

İlgilenilen değişkenin dağılımının başka değişkenlere bağlı olup olmadığının belirlenmesinde de bir değişken ile diğer bir değişkenin her bir seviyesinde dağılımını gösteren bir çapraz tablo oluşturulabilir (Howell, 2010).

1.3.1. İki Yönlü Çapraz Tablolar

İstatistiksel analiz yöntemlerinde kategorik verilerin çapraz sınıflandırılmış tablolar halinde gösterimi oldukça sık kullanılmaktadır. Bu tabloların en basiti iki yönlü çapraz tablodur.

Satırda yer alan $i = 1, \dots, I$ düzeyli kategorik değişken A, sütunda yer alan $j = 1, \dots, J$ düzeyli kategorik değişken B için iki yönlü $I \times J$ boyutlu çapraz sınıflandırılmış tabloların genel gösterimi aşağıdaki tabloda gösterilmiştir.

Tablo 1.1. İki Yönlü $\boldsymbol{I \times J}$ Boyutlu Çapraz Tablo

A/B	B_1	B_2	...	B_J
A_1	n_{11}	n_{12}	...	n_{1J}
A_2	n_{21}	n_{22}	...	n_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_I	n_{I1}	n_{I2}	...	n_{IJ}

A değişkeninin $i.$ düzeyi ile B değişkeninin $j.$ düzeyinde yer alan hücrenin gözlenen frekans değeri n_{ij} olarak ifade edilir.

1.3.2. Üç Yönlü Çapraz Tablolar

Üç değişkenin yer aldığı 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 üç yönlü $I \times J \times K$ boyutlu bir çapraz tablo aşağıdaki tabloda gösterilmiştir.

Tablo 1.2. Üç Yönlü $I \times J \times K$ Boyutlu Çapraz Tablo

A_I	B_J	C_K			
		C_1	C_2	…	C_K
A_1	B_1	n_{111}	n_{112}	…	n_{11K}
	B_2	n_{121}	n_{122}	…	n_{12K}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	B_J	n_{1J1}	n_{1J2}	…	n_{1JK}
A_2	B_1	n_{211}	n_{212}	…	n_{21K}
	B_2	n_{221}	n_{222}	…	n_{22K}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	B_J	n_{2J1}	n_{2J2}	…	n_{2JK}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_I	B_1	n_{I11}	n_{I12}	…	n_{I1K}
	B_2	n_{I21}	n_{I22}	…	n_{I2K}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	B_J	n_{IJ1}	n_{IJ2}	…	n_{IJK}

1.4. ÇAPRAZ TABLOLARIN ANALİZİNDE KULLANILAN OLASILIK (ÖRNEKLEME) DAĞILIMLARI

Sürekli değişkenlerin analizinde kullanılan istatistiksel yöntemler kategorik verilerin analizinde kullanılamamaktadır. Sürekli değişkenlere uygulanan birçok analizde dağılım varsayımı bulunmaktadır. Örneğin istatistiksel sonuç çıkarımında regresyon modelleri ve ANOVA (Analysis of Variance) normal dağılım varsayımını gerektirirken kategorik verilerin analizinde bu tür varsayımlar aranmamaktadır. Binom, Poisson, Çok

Terimli ve Çarpım-Çok Terimli dağılımları çapraz tabloların analizinde en sık kullanılan dağılımlardır (Christensen, 1997). Bu bölümde ilgilenilen dağılımlar kısaca özetlenmiştir.

1.4.1. Binom Dağılım (İki Terimli Dağılım)

Olasılık dağılımları içerisinde en yaygın kullanılan dağılımlardan biridir. İki mümkün sonucu olan bir deney n kez tekrarlandığında istenilen olay X defa meydana geliyorsa bu olayın olasılığı Binom dağılımı ile ifade edilir.

Binom dağılımının varsayımları;

- Mümkün olan iki sonuç vardır. “Başarılı” ve “Başarısız” olarak tanımlanabilir.
- Bir olayın “Başarılı” olma olasılığı "π" ise “Başarısız” olma olasılığı "$1-\pi$" olarak tanımlanır.
- Her deneme birbirinden bağımsız olup bir denemenin sonucu diğer bir denemeyi etkilemez. Yani, birbirinden bağımsız n adet “Bernoulli” deneyinin bir araya gelmesi sonucu n denemedeki başarısı sayısı X binom değişkeni Binom dağılır.

Binom dağılımının olasılık fonksiyonu;

X rasgele değişkeninin n denemedeki başarı sayısı;

$$P(X=x)=f(x)=b(x;n;\pi)=\binom{n}{x}\pi^{x}(1-\pi)^{n-x}\ ,\ x=0,1,2,3,\dots,n \qquad (1.1)$$

olasılık fonksiyonu ile gösterilir. Burada n deney sayısı ve π başarı olasılığı olmak üzere Binom dağılımı iki parametreye sahiptir.

Binom dağılımının beklenen değeri ve varyansı sırasıyla;

$$E(X)=n\pi \qquad (1.2)$$

$$Var(X)=n\pi(1-\pi) \qquad (1.3)$$

şeklinde ifade edilir.

1.4.2. Poisson Dağılım

Poisson dağılımı, bir olayın belirli bir birim; zaman, hacim, uzunluk, alan vb.'de gelme sayısının olasılık dağılımını ifade eder. Belirli bir altgeçidin altından bir dakika içersinde geçen otomobil sayısı veya yarım saat içersinde müşteri hizmetlerine gelen aramaların sayısı Poisson rasgele değişkenine örnektir.

Poisson dağılımının olasılık fonksiyonu;

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0,1,2,..,n \tag{1.4}$$

olup, bir olayın ilgilenilen periyodunda ortalama meydana gelme sayısını ifade eden λ, dağılımın tek parametresidir. Poisson dağılımının beklenen değeri ve varyansı ise;

$$E(X) = Var(X) = \lambda \tag{1.5}$$

şeklinde ifade edilir ve birbirine eşittir.

1.4.3. Çok Terimli Dağılım (Multinomial Dağılım)

Çok terimli dağılım binom dağılımının genelleştirilmiş hali olup ikiden fazla (k tane) mümkün sonuçlar içeren deneyler için kullanılan kesikli bir olasılık dağılımdır.

$i = 1,2, \dots, k$ olmak üzere π_i her denemede $i.$ olayın meydana gelme olasılığıdır ve $\sum_{i=1}^{k} \mu_i = 1$ şeklinde gösterilir. n_i, i.mümkün sonucun kaç kez meydana geldiğini göstersin. Bu durumda;

$$(X_1, X_2, \dots, X_k) \sim Mult(n, \pi_1, \dots, \pi_k)$$

Dağılımın olasılık fonksiyonu; $x_i \geq 0$, $x_1 + \cdots + x_k = n$ koşulu altında

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k} \tag{1.6}$$

fonksiyonu ile gösterilir. Bu durumda $X_i \sim Bin(n, \pi_i)$ olup, X_i'lerin beklenen değer ve varyansı;

$$E(X_i) = n\pi_i \tag{1.7}$$

$$Var(X_i) = n\pi_i(1 - \pi_i) \tag{1.8}$$

şeklinde ifade edilir. Bu tez çalışmasında çok terimli dağılımdan faydalanılmıştır.

İKİNCİ BÖLÜM

LOGARİTMİK DOĞRUSAL MODELLER

İstatistiksel analiz yöntemlerinin en önemli aşamalarından biri değişken yapılarına göre uygulanacak analiz yönteminin belirlenmesidir. Kategorik verilerin analizinde logaritmik doğrusal modeller sıklıkla kullanılır. Logaritmik doğrusal modeller çok yönlü çapraz tablolarda, değişkenler arasındaki ilişkileri test etmek üzere kullanılan istatistiksel bir teknik olup yığına ilişkin tahminlerde bulunulmasına yardımcı olur. Genellikle iki veya daha fazla değişkenin (çok yönlü çapraz tabloların) analizinde kullanıldığından "Çok-Yönlü Frekans Analizi" olarak da adlandırılır.

Logaritmik doğrusal analiz, iki veya daha fazla kategorik değişken arasındaki koşullu ilişkinin incelendiği iki yönlü çapraz tablosunun bir açılımıdır. Logaritmik doğrusal modeller çok yönlü çapraz tablolarda bağımlılık yapılarının analizinde kullanılan istatistiksel bir metottur (Brzezinska, 2013). Bu modeller hücrelerdeki frekansların, değişkenlerin düzeylerine ne kadar bağlı olduklarını ortaya koyan modeller olup Poisson ve Çokterimli dağılımlı veri yapısına uygundur. Çapraz tablolardaki sayım verilerinin Poisson dağıldığı durumda hücrelerdeki frekansların birbirinden bağımsız olduğu kabul edilirken, çok terimli dağılım durumunda ise hücrelerin birbirinden bağımsız olmadığı kabul edilmektedir (Çilan, 2013). Logaritmik doğrusal modeller Poisson dağılımına sahip veri için genelleştirilmiş doğrusal modellerin özel bir durumudur.

Logaritmik doğrusal analiz karmaşık yapılı çok yönlü tabloların analizine sistematik bir yaklaşım getirir. İlgilenilen etkilerin büyüklüğünün tahmin edilmesine ve buna bağlı olarak incelenecek farklı etkilerin göreli öneminin belirlenmesine olanak sağlar.

Logaritmik doğrusal modellerde incelenen tüm değişkenler yanıt değişkenleri olarak ele alınır. 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). 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 kullanılmalıdır.

Pearson χ^2 test istatistiği veya olabilirlik oran istatistiği L^2 (Likelihood Ratio Statistics), sadece iki kategorik değişkenden oluşan iki yönlü çapraz sınıflandırılmış tabloların analizinde kullanılır. Buna karşılık üç veya daha fazla değişkenden oluşan çok-boyutlu tabloların analizinde ise logaritmik doğrusal modeller kullanılmaktadır.

Logaritmik doğrusal model analizi kesikli çok değişkenli analizler arasında önemli bir yere sahiptir. İki kategorik değişkenin ele alındığı iki yönlü çapraz tablolarda sadece bağımsızlık test edilirken, üç veya daha fazla sayıda değişkenin oluşturduğu çapraz tabloların analizinde temel ve etkileşim etkilerine ait hipotezler de test edilir.

Uygun (1990), 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 toplanmıştır.

2.1. İKİ YÖNLÜ ÇAPRAZ TABLOLARDA LOGARİTMİK DOĞRUSAL ANALİZ

Satırda yer alan $i = 1, \ldots, I$ düzeyli kategorik değişken A ile sütunda yer alan $j = 1, \ldots, J$ düzeyli kategorik değişken B için iki yönlü $I \times J$ boyutlu çapraz sınıflandırılmış tabloların genel gösterimi aşağıda verilmiştir.

Tablo 2.1. A ve B Değişkenlerine ait $I \times J$ Boyutlu Çapraz Tablo

A/B	B_1	B_2	$\cdots$	B_J
A_1	n_{11}	n_{12}	$\cdots$	n_{1J}
A_2	n_{21}	n_{22}	$\cdots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}

n_{ij}, A değişkeninin i. düzeyi ile B değişkeninin j. düzeyinde yer alan hücrenin gözlenen frekans değerini ifade eder. 2×2'lik bir çapraz tablo için A değişkeninin i. satıra ve B değişkeninin j. sütuna düşme olasılığı π_{ij}, A ve B değişkenleri bağımsız olduğunda;

$$\pi_{ij} = \pi_{i.}\pi_{.j} \ ; i = 1,2, \dots, I \text{ ve } j = 1,2, \dots, J \tag{2.1}$$

şeklinde ifade edilir. Eşitlik 2.1'de çarpımsal olarak ifade edilen modelin doğal logaritması alınarak, logaritmik doğrusal modeller,

$$log\pi_{ij} = log\pi_{i.} + log\pi_{.j} \tag{2.2}$$

olarak gösterilir. Beklenen sıklıkları m_{ij} ile göstermek üzere (2.1)'de çarpımsal olarak ifade edilen modeldeki beklenen sıklıklar,

$$m_{ij} = n\pi_{ij} = n\pi_{i.}\pi_{.j} \tag{2.3}$$

olarak gösterilir ve modelin doğal logaritması alındığında,

$$logm_{ij} = logn + log\pi_{i.} + log\pi_{.j} \tag{2.4}$$

şeklinde ifade edilir.

Beklenen sıklıklar aynı zamanda,

$$m_{i.} = n\pi_{i.} \quad \text{ve} \quad m_{.j} = n\pi_{.j} \tag{2.5}$$

olarak verilir ve doğal logaritması alınarak,

$$logm_{i.} = logn + log\pi_{i.}$$

$$logm_{.j} = logn + log\pi_{.j} \qquad (2.6)$$

eşitliği elde edilir. Eşitlik 2.6'da, verilen eşitlikler,

$$log\pi_{i.} = logm_{i.} - logn$$

$$log\pi_{.j} = logm_{.j} - logn \qquad (2.7)$$

şeklinde ifade edilebilir. Bu eşitlikler Eşitlik 2.4'te yazıldığında,

$$logm_{ij} = logm_{i.} + logm_{.j} - logn \qquad (2.8)$$

eşitliği bulunur. Eşitlik 2.8'de verilen ifadenin $i = 1,2,\dots,I$ ve $j = 1,2,\dots,J$ üzerinden toplam alınırsa,

$$\sum_{i=1}^{I}\sum_{j=1}^{J} logm_{ij} = J\sum_{i=1}^{I} logm_{i.} + I\sum_{j=1}^{J} logm_{.j} - IJlogn \qquad (2.9)$$

elde edilir.

$$\lambda_i = logm_{i.} - \frac{1}{I}\sum_{i=1}^{I} logm_{ij}$$

$$\lambda_j = logm_{.j} - \frac{1}{J}\sum_{j=1}^{J} logm_{ij}$$

$$\lambda_0 = \frac{1}{I}\sum_{i=1}^{I} logm_{i.} + \frac{1}{J}\sum_{j=1}^{J} logm_{.j} + logn \qquad (2.10)$$

olarak tanımlandığında, Eşitlik 2.9'da, λ_i, λ_j ve λ_0 parametreleri cinsinden,

$$\mathrm{log}m_{ij} = \lambda_0 + \lambda_i^A + \lambda_j^B \qquad (2.11)$$

şeklinde ifade edilir. Bu model A ve B gibi iki değişken arasında etkileşim etkisinin olmadığı bağımsız logaritmik doğrusal model olup, doymamış logaritmik doğrusal modeldir.

Doymuş logaritmik doğrusal model, iki değişken arasında etkileşim etkisinin olduğu modeldir ve

$$\log m_{ij} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \tag{2.12}$$

şeklinde gösterilir. Bu modelde;

$\log m_{ij}$: Beklenen frekansların doğal logaritması,

λ_0 : Sabit (Genel Ortalama)

λ_i^A : A değişkeninin i. düzey ana etkisi

λ_i^B : B değişkeninin j. düzey ana etkisi,

λ_{ij}^{AB}: A değişkeninin i. düzeyinde, B değişkeninin j. düzeyinde ikinci dereceden etkileşim etkisi,

olarak tanımlanmıştır.

Eşitlik 2.12'de verilen logaritmik doğrusal modele ait parametrelerin aşağıda tanımlanan,

$$\sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0 \tag{2.13}$$

kısıtları sağlaması gerekir. λ_{ij}^{AB} parametresinin 0 değerini alması, A değişkeni ile B değişkeninin bağımsız olmasını yani birlikte değişim ve etkileşimin olmadığı, Eşitlik 2.11'deki doymamış (unsaturated) modeli ifade eder.

k değişken sayısı olmak üzere doymuş bir modelde, λ_0 sabit parametresi dışında 2^k tane parametre vardır. İki değişkenli logaritmik doğrusal model için model parametreleri ve serbestlik dereceleri Tablo 2.2'de verilmiştir (Demirhan, 2004).

Tablo 2.2. İki Yönlü Logaritmik Doğrusal Model Parametre ve Serbestlik Dereceleri

Parametre	Serbestlik Derecesi
$\boldsymbol{\lambda_0}$	1
$\boldsymbol{\lambda_i^A}$	$I-1$
$\boldsymbol{\lambda_j^B}$	$J-1$
$\boldsymbol{\lambda_{ij}^A}$	$(I-1)\,(J-1)$
Toplam	IJ

İki yönlü çapraz tablolarda iki değişken arasında etkileşim olup olmadığı,

$$H_0: \lambda_{ij}^{AB} = 0 \; (Etkileşim\; yoktur)$$

$$H_1: \lambda_{ij}^{AB} \neq 0 \; (Etkileşim\; vardır) \tag{2.14}$$

hipotezleri ile verilir. H_0, yokluk hipotezi geçerli ise, bu değişkenlere ait model,

$$logm_{ij} = \lambda_0 + \lambda_i^A + \lambda_i^B \tag{2.15}$$

ile verilen bağımsız model olacaktır.

Eşitlik 2.14'de kurulan hipotezlerin test edilmesinde Pearson χ^2 test istatistiği ve Log-olabilirlik oran istatistiği L^2 kullanılmaktadır. Pearson χ^2 test istatistiği; n_{ij} çapraz tabloda yer alan her bir hücrenin gözlenen frekansı, m_{ij} çapraz tabloda yer alan her bir hücrenin beklenen frekansı olmak üzere iki boyutlu ve üç boyutlu çapraz tabloları için sırasıyla;

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij}-m_{ij})^2}{m_{ij}} \sim \chi_{sd}^2 , \qquad sd = (I-1)(J-1) \tag{2.16}$$

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(n_{ijk}-m_{ijk})^2}{m_{ijk}} \sim \chi_{sd}^2 , \quad sd = (I-1)(J-1)(K-1) \tag{2.17}$$

eşitlikleri ile hesaplanır.

χ^2 test istatistiği gözlenen frekanslar ile beklenen frekanslar arasındaki farkın anlamlılığının test edilmesi temeline dayanır ve serbestlik derecesi (sd) ile karakterize edilen sürekli bir dağılımdır. Serbestlik derecesi arttıkça normal dağılıma yakınsar. χ^2 dağılımın ortalaması sd'ye, varyansı ise sd'nin 2 katına eşittir. χ^2 testi, değişkenler arasında ilişki olup olmadığını belirler ancak ilişkinin yönü ve büyüklüğü hakkında bilgi vermez.

Olabilirlik oran istatistiği L^2, iki boyutlu ve üç boyutlu çapraz tabloları için sırasıyla;

$$L^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log(\frac{n_{ij}}{m_{ij}}) \tag{2.18}$$

$$L^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log(\frac{n_{ijk}}{m_{ijk}}) \tag{2.19}$$

eşitlikleri ile bulunur. L^2 test istatistiği, iki ve üç yönlü çapraz tabloları için sırasıyla $(I-1)\,(J-1)$ ve $(I-1)\,(J-1)(K-1)$ serbestlik dereceleri altında yaklaşık χ^2 dağılır.

2.2. ÜÇ YÖNLÜ ÇAPRAZ TABLOLARDA LOGARİTMİK DOĞRUSAL ANALİZ

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 (satır), j (sütun) ve k (tabaka)'yı ifade etmek üzere (i,j,k) gözesinin gözlenen frekansı n_{ijk} ile gösterilir. A değişkeninin i'inci satır, B değişkeninin j'inci sütun ve C değişkeninin k'ıncı tabakasına düşme olasılığını π_{ijk} verir. n_{ijk}'ların olasılık dağılımı A, B ve C değişkenlerinin ortak olasılık dağılımıdır.

Bu durumda 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,

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}\ . \tag{2.20}$$

ve bu model “Doymuş Logaritmik Doğrusal Model” olarak adlandırılır. Bu model, aynı zamanda “Hiyerarşik Logaritmik Doğrusal Model” olarak adlandırılır ve λ_0 genel ortalama olmak üzere, değişkenlere ait ana etkileri, bütün olası ikili etkileşim etkilerini ve üçlü etkileşim etkilerini içeren parametreleri içerir.

$\log(m_{ijk})$: Beklenen frekansların doğal logaritması, olmak üzere modeldeki parametreler,

λ_0 : Sabit (Genel Ortalama),

λ_i^A : A değişkeninin i. düzey ana etkisi,

λ_j^B : B değişkeninin j. düzey ana etkisi,

λ_k^C : C değişkeninin k. düzey ana etkisi,

λ_{ij}^{AB} : A’nın i. düzeyinde (λ_i^A), B’nin j. düzeyinde (λ_j^B) ikinci dereceden etkileşim etkileri,

λ_{ik}^{AC} : A’nın i. düzeyinde (λ_i^A) ve C’nin k. düzeyinde (λ_k^C) ikinci dereceden etkileşim etkileri,

λ_{jk}^{BC} : B’nin j. düzeyinde (λ_j^B) ve C’nin k. düzeyinde (λ_k^C) ikinci dereceden etkileşim etkileri,

λ_{ijk}^{ABC} : A’nın i. düzeyinde (λ_i^A), B’nin j. düzeyinde (λ_j^B) ve C’nin k. düzeyinde (λ_k^C) üçüncü dereceden etkileşim etkileridir.

Doymuş modele ait parametreler ve serbestlik dereceleri aşağıdaki tabloda verilmiştir.

Tablo 2.3. Üç Yönlü Logaritmik Doğrusal Model Parametre ve Serbestlik Dereceleri

Parametreler	Serbestlik Dereceleri
$\boldsymbol{\lambda_0}$	1
$\boldsymbol{\lambda_i^A}$	$(I-1)$
$\boldsymbol{\lambda_j^B}$	$(J-1)$
$\boldsymbol{\lambda_k^C}$	$(K-1)$
$\boldsymbol{\lambda_{ij}^{AB}}$	$(I-1)\,(J-1)$
$\boldsymbol{\lambda_{ik}^{AC}}$	$(I-1)\,(K-1)$
$\boldsymbol{\lambda_{jk}^{BC}}$	$(J-1)\,(K-1)$
$\boldsymbol{\lambda_{ijk}^{ABC}}$	$(I-1)(J-1)\,(K-1)$

Üç yönlü çapraz tablolara ait toplam 9 adet logaritmik doğrusal model ise aşağıdaki tabloda özetlenmiştir (Andersen, 1990).

Tablo 2.4. Üç Yönlü Çapraz Tablolarda Olası Hiyerarşik Logaritmik Doğrusal Modeller

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

Üç yönlü çapraz tabloda yer alan olası 9 adet logaritmik doğrusal modelin yorumu aşağıda belirtilmiştir.

1. $M^{(0)}$ modeli bütün değişkenlerin birbirinden bağımsız olduğu tam bağımsızlık modelini ifade eder. Model sadece ana etkileri ($\lambda_i^A, \lambda_j^B, \lambda_k^C$) ve genel ortalamayı ($\lambda_0$) içerir.

2. $M^{(1)}, M^{(2)}, M^{(3)}$ modelleri bir değişken çiftinin bağımlı, diğer değişkenlerin ise birbirinden bağımsız olduğunu ifade eden modellerdir. Örneğin;

$M^{(2)}$: $\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC}$ logaritmik doğrusal modeline göre B değişkeninin $j.$ düzeyi verildiğinde A ve C değişkenleri koşullu bağımlıdır. Bu bağımlılığı λ_{ik}^{AC} parametresi ifade eder.

3. $M^{(4)}, M^{(5)}, M^{(6)}$ modelleri tüm olası ikili etkileşim etkilerinden birinin model dışı bırakıldığı ve üçlü etkileşim etkilerinin olmadığı modellerdir. Örneğin;

M$^{(5)}$: $\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$ modeline göre A değişkeninin $i.$ düzeyi verildiğinde B ve C değişkenlerinin birbirinden bağımsız olduğunu yani aralarında ilişkinin olmadığını gösterir. Bu bağımlılığı gösteren λ_{jk}^{BC} parametresi modelde yer almaz.

4. $M^{(7)}$: $\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$ logaritmik doğrusal modeli tüm ikili etkileşimleri içeren ancak üçlü etkileşimleri içermeyen bir modeldir. Bu modele göre C değişkeninin $k.$ düzeyi verildiğinde A ve B değişkenleri (λ_{ij}^{AB}), B değişkeninin $j.$ düzeyi verildiğinde A ile C değişkenleri (λ_{ik}^{AC}), A değişkeninin $i.$ düzeyi verildiğinde B ile C değişkenleri (λ_{jk}^{BC}) koşullu bağımlıdır.

5. $M^{(8)}$ 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 gerekir. Yani bu model en karmaşık etkileşim yapısından en basit etkilere kadar ifade edilebilir.

Hiyerarşik olmayan logaritmik doğrusal modeller bazı istatistiksel paket programları ile test edilememektedir. Çünkü hiyerarşik olmayan modeller arasından seçim için bir istatistiksel süreç sağlanamamaktadır. Bunun sebebi ise bir etkileşim teriminin anlamlı olmadığı durumlarda, daha yüksek dereceli etkileşimlerin anlamlı olup olmadığının yorumunun anlam taşımamasındandır (Oğuzlar, 2004:237).

Üç yönlü çapraz tablolar için ifade edilen logaritmik doğrusal modeller, çok yönlü çapraz tablolar ($k \geq 4$) için de genişletilebilir. Ancak değişken sayısı arttıkça, parametre sayısı artmakta, ilişki ve etkileşim terimlerini açıklamak karmaşık, kolay yorumlanamayan bir hal almaktadır. Bu nedenle bu çalışmada üç yönlü çapraz tablolar için hiyerarşik logaritmik doğrusal modeller incelenmeye çalışılmış ve ayrıntılarıyla verilmiştir.

2.2.1. Üç Yönlü Hiyerarşik Logaritmik Doğrusal Modellerde Test Edilecek Hipotezler

Hiyerarşik logaritmik doğrusal modellerde ana etkiler ve etkileşim parametrelerinin test edilmesine ilişkin Tablo 2.4'te verilen hiyerarşik logaritmik doğrusal modellere ait test edilecek hipotezler aşağıda özetlenmiştir. Tam bağımsızlığı ifade eden $M^{(0)}$ modeli için test edilecek hipotez,

$$H_0 : \lambda_{ijk}^{ABC} = \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = 0 \text{ , bütün } i, j \; ve \; k \; için \quad (2.21)$$

şeklinde olup, bütün değişkenlerin birbirinden bağımsız olduğunu ifade eder. H_0 yokluk hipotezi red edilemediğinde uygun model;

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C \quad (2.22)$$

şeklinde olacaktır. Üç değişkenli etkileşim etkisinin olmadığı $M^{(7)}$ modeli için test edilecek hipotez,

$$H_0 : \lambda_{ijk}^{ABC} = 0 \text{ , } \quad \text{bütün } i, j \; ve \; k \; için \quad (2.23)$$

şeklinde olup, A, B ve C değişkenlerinin etkileşim etkisinin istatistiksel olarak anlamlı olmadığını ifade eden doymuş hiyerarşik logaritmik doğrusal modelin alternatifine karşı sınandığı önsavdır. H_0 yokluk hipotezinin red edilemediği durumda uygun model;

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (2.24)$$

şeklinde olacaktır. İki değişkenli etkileşim etkilerinden ikisinin olmadığı $M^{(3)}$ modeli için test edilecek hipotez,

$$H_0 : \lambda_{ijk}^{ABC} = \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = 0, \quad \text{bütün } i, j \ ve \ k \ için \tag{2.25}$$

şeklinde olup, H_0 yokluk hipotezinin red edilemediği durumda uygun model,

$$\log\left(m_{ijk}\right) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC} \tag{2.26}$$

şeklinde olacaktır. İki değişkenli etkileşim etkilerinden birinin olmadığı $M^{(5)}$ modeli için test edilecek hipotez ise,

$$H_0 : \lambda_{ijk}^{ABC} = \lambda_{jk}^{BC} = 0, \quad \text{bütün } i, j \ ve \ k \ için \tag{2.27}$$

şeklinde olup, A değişkeninin $i.$ düzeyi verildiğinde, B ve C değişkenlerinin etkileşim etkisinin olmadığı önsavdır. H_0 yokluk hipotezinin red edilemediği durumda verilere en uygun model,

$$\log\left(m_{ijk}\right) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} \tag{2.28}$$

şeklinde olacaktır.

2.3. ÇAPRAZ TABLOLARDA İLİŞKİ ÖLÇÜTLERİ

Bu bölümdeki ilişki ölçütleri Agresti (1996)’den yararlanılarak açıklanmıştır.

2.3.1. Oranlar Farkı

A ve B iki düzeyli (başarılı/başarısız) değişkenler olsun. 1. satır için başarının olasılığı π_1 ise başarısız olma olasılığı $1 - \pi_1$’dir. Aynı şekilde 2. satır için başarının olasılığı π_2 başarısızlığın olasılığı $1 - \pi_2$ olur. Oranlar farkı $\pi_1 - \pi_2$ olarak ifade edilir ve bu fark -1 ile +1 değerleri arasında yer alır. $\pi_1 = \pi_2$ grupların bağımsızlığını ifade eder.

Başarının örnek oranları 1. satır için p_1, 2. satır için p_2 olarak gösterilsin. Bu durumda örnek oranları $p_1 = n_{11}/n_{1.}$; $p_2 = n_{21}/n_{2.}$ olup, satır toplamları sırası ile $n_{1.}$ ve $n_{2.}$’dir. Örnek oranları $p_1 - p_2$ olmak üzere bu farkın tahmini standart hatası,

$$\hat{\sigma}(p_1 - p_2) = \hat{\sigma}_{(p_1-p_2)} = \sqrt{\frac{p_1(1-p_1)}{n_{1.}} + \frac{p_2(1-p_2)}{n_{2.}}} \qquad (2.29)$$

şeklinde gösterilir ve

$\pi_1 - \pi_2$ için $\%(1-\alpha) \times 100$ için Güven Aralığı,

$$(p_1 - p_2) \pm z_{\alpha/2}\, \hat{\sigma}(p_1 - p_2) \qquad (2.30)$$

ile ifade edilir. Burada, $z_{\alpha/2}$ standart normal dağılım tablosundan sağ kuyruk alanı $\frac{\alpha}{2}$ yapan kritik değerdir.

2.3.2. Göreli Risk (Relative Risk)

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

$$GR = \frac{\pi_1}{\pi_2} \qquad (2.31)$$

şeklinde hesaplanır.

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

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

$\log\left(\frac{\pi_1}{\pi_2}\right)$ için % $(1-\alpha) \times 100$ Güven Aralığı;

$$\log\left(\frac{p_1}{p_2}\right) \pm z_{\alpha/2} \cdot \sqrt{\frac{1-p_1}{n_1 . p_1} + \frac{1-p_2}{n_2 . p_2}} \qquad (2.32)$$

ile hesaplanır. Bu güven aralığında varyans bilinmediğinden tahmin edicisi kullanılır. $\log\left(\frac{\pi_1}{\pi_2}\right)$ için güven aralığı $(AS, \text{Ü}S)$ olarak alınırsa, $\frac{\pi_1}{\pi_2}$ için % $(1-\alpha) \times 100$ güven aralığı, $(e^{AS}, e^{\text{Ü}S})$ olacaktır. (Dağalp, 2018, Tang, He, & Tu, 2012).

2.3.3. Odds Oranı (Çapraz Çarpım Oranı)

Çapraz çarpım 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,

$$odds_1 = \frac{\pi_1}{(1-\pi_1)} \tag{2.33}$$

$$odds_2 = \frac{\pi_2}{(1-\pi_2)} \tag{2.34}$$

şeklinde tanımlanır. Örneğin; $\pi = 0.75$ ise odds değeri, $\text{odds} = 0.75/0.25 = 3$ olur. Bu değer başarının gerçekleşme olasılığının başarısızlığa göre 3 kat daha fazla olduğunu gösterir. 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 eder. Ayrıca herhangi bir olayın meydana gelmesinin odds'u biliniyorsa olayın gerçekleşme olasılığı;

$$\pi = \frac{odds}{odds+1} \tag{2.35}$$

şeklinde ifade edilir. İki ayrı odds değerinin ya da koşullu odds değerinin birbirine oranına "Çapraz Çarpım Oranı (Cross Product Ratio)" veya "Odds Oranı" denir. "θ" veya "OR" ile tanımlanır.

$$\theta = \frac{odds_1}{odds_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} \tag{2.36}$$

şeklinde hesaplanır. Eşitlik 2.31 ve 2.36'dan anlaşılacağı üzere göreli risk iki olasılığın birbirine oranı iken odds oranı θ, iki odds'un birbirine oranıdır.

2.3.3.1. Odds Oranının Özellikleri

Odds oranının özellikleri aşağıdaki gibi özetlenebilir:

- Odds oranı negatif olmayan herhangi bir tamsayıya eşittir.
- Odds oranının değeri satır ve sütun değerlerinin yer değiştirmesiyle değişmez, yani simetriktir (invariant).
- 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 eder.
- Odds oranı 1'den uzaklaştıkça değişkenler arasındaki ilişki kuvvetlenir.
- Odds oranı $1 < \theta < \infty$ olduğunda 1. satırdaki başarının odds'u 2. satırdaki başarının odds'undan daha yüksektir. Örneğin; $\theta = 4$ olduğunda 1. satırdaki başarının odds'u 2. satırdaki başarının odds değerinin 4 katıdır ($\pi_1 > \pi_2$).
- Odds oranı $0 < \theta < 1$ olduğunda ise 1. satırdaki başarının odds'u 2. satırdaki başarının odds'undan daha düşüktür ($\pi_1 < \pi_2$).
- Odds oranının örneklem genişliği ile ilişkisi yoktur. Örneklem genişliği, çapraz tablonun tüm gözelerinde aynı oranda artarsa odds oranı aynı kalır (Liao, 1994 : 48).
- Odds oranının en çok olabilirlik tahmin edicisi $\hat{\theta}$,

$$\hat{\theta} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \tag{2.37}$$

şeklinde ifade edilir.

2.3.4. Odds Oranı ve Logaritmik Odds Oranı

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 örneklemlerde ortalaması log θ ve asimptotik standart hata (ASE) olarak adlandırılan bir standart sapması ile normal dağılıma yakınsar. 2×2 lik bir çapraz tablo için standart hata,

$$\text{ASE}(\log \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \tag{2.38}$$

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

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

$$\log(\hat{\theta}) \pm z_{\alpha/2}\text{ASE}(\log \hat{\theta}) \tag{2.39}$$

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

$$AS = \log(\hat{\theta}) - z_{\alpha/2}\text{ASE}(\log \hat{\theta}) \tag{2.40}$$

$$\text{Ü}S = \log(\hat{\theta}) + z_{\alpha/2}\text{ASE}(\log \hat{\theta}) \tag{2.41}$$

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

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.

Düzeltilmiş Çapraz Çarpım Oranı

Çapraz tablolarda herhangi bir göze sıklığı $n_{ij} = 0$ ise örnek çapraz çarpım oranı da 0 veya +∞ olur. Bu durumda her hücreye 0.5 eklenerek '$\tilde{\theta}$' ile gösterilen düzeltilmiş çapraz çarpım oranı,

$$\tilde{\theta} = \frac{(n_{11}+0.5)(n_{22}+0.5)}{(n_{12}+0.5)(n_{21}+0.5)} \tag{2.42}$$

olarak hesaplanır.

2.3.5. Odds Oranı (OR) ve Göreli Risk (GR) Arasındaki İlişki

Odds oranı ve göreli risk her iki grup için başarı oranının sıfıra yakın olması durumda benzer değerler alır.

$$\text{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \text{GR} \times (\frac{1-p_2}{1-p_1}) \quad (2.43)$$

Odds oranı ve göreli risk arasındaki bu ilişki kullanışlıdır. Bazı veri setleri için göreli riskin hesaplanması mümkün değildir ancak odds oranı hesaplanabilir ve bunun sonucunda göreli riski yaklaşık olarak bulunabilir.

2.4. UYGUN MODELİN SEÇİMİ

Verilen bir çapraz tabloya ait logaritmik doğrusal analiz yapılırken, elde edilen modelin ilgili yığın için en uygun model olması gerekmektedir. Model seçimi yapılırken iki tip hata karşımıza çıkabilir. İlk olarak seçilen model gereğinden fazla parametre içerebilir ya da seçilen modelde yer alması gereken parametreler modelde yer almayabilir. Her iki durumda da veri ile model arasında bir uyumsuzluk problemi ortaya çıkabilir. Bu nedenle en uygun model aşağıdaki kriterleri taşımalıdır (Becanım, 2006) :

- En az parametreye sahip olmalıdır.
- Ele alınan uyum iyiliği testlerine göre anlamlı olmalıdır.

Araştırmacının kolayca yorumlayabileceği model, ele alınan yığın için en uygun modeldir.

Birden fazla modelin anlamlı çıkması halinde, hangi logaritmik doğrusal modelin seçilen değişkenler arasındaki ilişkiyi doğru ifade ettiğini belirlemede Pearson χ^2ve Log-olabilirlik oran istatistiği L^2 gibi bazı uyum iyiliği testleri kullanılmaktadır. Ancak uyum iyiliği testlerinden en yaygın kullanılanları olabilirlik oran istatistiği L^2 'dir (Karabulut, 1998: s.11; Agresti, 1990: s.174).

L^2 istatistiğinin, Pearson χ^2 istatistiğine tercih edilmesinin iki temel nedeni vardır. Birincisi beklenen frekansların en çok olabilirlik yöntemi ile tahmin edilmesi, ikincisi ise L^2 istatistiğinin çok yönlü frekans tablolarında koşullu bağımsızlıkların daha güçlü test edilmesini sağlayacak biçimde parçalara ayrılabiliyor olmasıdır (Knoke & Burke, 1980:30).

Araştırmacı, birden fazla aşamalı logaritmik doğrusal modele sahip ise bu modeller içinden en uygun modele karar verirken L^2 test istatistiğini kullanmalıdır.

Örneğin; M_1 tüm değişkenlerin birbirinden bağımsız olduğunu ifade eden tam bağımsızlık modeli [A][B][C] ve M_2 bir değişken çiftinin bağımlı, diğer değişkenlerin ise birbirinden bağımsız olduğunu ifade eden modellerden [BC] [A] hiyerarşik logaritmik doğrusal modeli olsun. Bu iki modelden hangisinin yığın için en uygun model olduğuna karar verilirken M_1 ve M_2 modelleri için hesaplanan L^2 olabilirlik oran test istatistikleri farkları alınarak karar verilir. Bu test istatistiğine koşullu (şartlı) olabilirlik oran test istatistiği adı verilir ve

$$L^2(M_1/M_2) = L^2(M_1) - L^2(M_2) \tag{2.44}$$

şeklinde ifade edilir. Koşullu olabilirlik oran test istatistiği, $sd_1 - sd_2$ serbestlik derecesi altında asimptotik olarak χ^2 dağılımına sahiptir. Bu nedenle ele alınan modellere ait L^2 değerlerinin farkı $sd(M_1) - sd(M_2)$ serbestlik derecesindeki χ^2 değeriyle karşılaştırıldığında $p < \alpha$ ise modeller arasında fark olduğu ve test edilen değişkenler arasında etkileşim bulunmadığını ifade eden, $H_0 : \lambda_{ij}^{AB} = 0$ yokluk hipotezinin reddedilmesi gerektiği söylenir. Bu durumda uygun model M_2 modeli olur. Eğer aralarındaki farklılık anlamlı değilse uygun model M_1 modeli olacaktır.

2.4.1. Uygun Modelin Seçiminde Kullanılan Adımsal Yöntemler

Bir yığın için onu temsil edecek en uygun logaritmik doğrusal model seçilirken ileriye dönük seçim yöntemi veya geriye dönük eleme yöntemi uygulanır. Bu yöntemleri üç değişkenli logaritmik doğrusal modeller üzerinde aşağıdaki gibi inceleyebiliriz.

İleriye dönük seçim yönteminde modelleme işlemi $M^{(0)}$ tam bağımsızlık modeli $\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C$, ile başlar. Verilere en uygun model seçilinceye kadar kendisinden bir fazla parametre içeren modellerle karşılaştırılarak eklenen parametrenin anlamlı olmadığını söyleyen H_0 yokluk hipotezi red edilemeyene kadar tekrar edilir. H_0 yokluk hipotezinin red edilememesinden önceki model, yığın için en uygun modeldir.

Geriye dönük eleme yönteminde modelleme işlemi değişkenler arasındaki olası tüm etkileşimleri içeren $M^{(8)}$, $\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$ doymuş hiyerarşik logaritmik doğrusal modeli ile başlar. Olası tüm etkileşimleri içeren doymuş modelden veri örneklerine en uygun daha basit bir model bulabilmek için istatistiksel olarak anlamlı bulunmayan parametreler modelden elenerek modelden çıkarılan parametrenin anlamlı olmadığını ifade eden H_0 yokluk hipotezi red edilinceye kadar devam eder. H_0 yokluk hipotezinin red edilmesi modelden daha fazla parametre elenemeyeceği anlamına gelir ve son adımda elde edilen model, yığın için geriye dönük eleme (stepwise) ile elde edilen verilere en uygun model olur.

2.4.2. Uygun Modelin Seçiminde Kullanılan Kriterler

Bir yığın için en uygun modelin seçiminde birden fazla model olması durumunda bir takım ölçütler kullanılmakta ve modelin yorumlanmasına yardımcı olmaktadır. Bunlardan birincisi s; örneklem genişliğidir. Burada N yığın çapı, L^2 olabilirlik oran istatistiği olmak üzere, örneklem genişliği s,

$$s = \frac{L^2}{N} \qquad (2.45)$$

şeklinde hesaplanır.

Örneklem genişliğinden yola çıkarak model seçiminde kullanılabilecek kriterlerden birisi $w = \sqrt{s}$'dir. Ancak sadece hesaplanan w değerine bakılarak model seçimine karar vermek doğru olmayabilir. Bu nedenle model seçimini diğer kriterlerle desteklemek faydalı olacaktır.

Uygun model seçiminin karar verilmesinde kullanılan bir diğer ölçüt,

$$R^2 = \frac{L^2(M_0) - L^2(M)}{L^2(M_0)} \qquad (2.46)$$

olup hiyerarşik logaritmik doğrusal modellerin seçiminde kullanılmaktadır. R^2, 0 ile 1 arasında değerler alan bir ölçüttür. Bu kritere göre R^2 değeri 1'e yaklaştıkça ele alınan bir sonraki model yani M modeli, 0'a yaklaştıkça M_0 modeli kabul edilir. Ancak R^2'nin hesabında yukarıda verilen eşitlikte görüldüğü üzere modeldeki parametre sayısının yer almaması araştırmacıyı yanılgıya düşürebilir. Bu nedenle R^2'ye alternatif bir ölçüt,

$$\delta = \frac{L^2(M_0)/sd(M_0) - L^2(M)/sd(M)}{L^2(M_0)/sd(M_0)} \qquad (2.47)$$

olarak sunulmuştur (Becanım, 2006:26).

δ değeri R^2'ye benzer şekilde yorumlanır. Alabileceği en yüksek değer 1'dir ve R^2'den küçüktür. Aynı anda aynı R^2 değerlerine sahip iki model varsa bu modeller içerisinde δ değeri büyük olan model o yığın için en uygun model olarak ele alınır (Becanım, 2006:26).

En uygun model seçiminde kullanılan diğer bir kriter, Akaike bilgi kriteri (Akaike's Information Criterion)'dir. Akaike Bilgi Kriteri, yığın için en uygun model ele alınırken, birden fazla hiyerarşik model varsa, her bir model için $L^2(M)$ farkları kullanılarak mevcut modeller, kendisinden daha fazla parametre içeren hiyerarşik modele karşı test edilir. Akaike Bilgi Kriteri, AIC ile gösterilir ve,

$$AIC = L^2(M) - [q - 2sd(M)] \tag{2.48}$$

şeklinde hesaplanır. Bu eşitlikte;

$L^2(M)$: M modelini kendisinden daha fazla parametre içeren modele karşı test eden olabilirlik oran istatistiği,

q : kendisinden daha fazla parametre içeren modele ait serbestlik derecesi,

$sd(M)$: M modeli için serbestlik derecesi,

olarak tanımlanır.

ÜÇÜNCÜ BÖLÜM

İNTİHAR OLASILIĞI ÖLÇEĞİ VE SOSYO-DEMOGROFİK DEĞİŞKENLERİN LOGARİTMİK DOĞRUSAL MODELLER YARDIMI İLE İNCELENMESİ

3.1. ARAŞTIRMANIN AMACI, KAPSAMI VE YÖNTEMİ

3.1.1. Amaç

Bu çalışmanın amacı lise ve üniversite öğrencilerine ait intihar olasılığı ölçeği, problem çözme envanteri, cinsiyet, sosyo-ekonomik düzey, öğrenim düzeyi gibi çeşitli sosyo-demogrofik değişkenler arasındaki ilişkilerin logaritmik doğrusal modeller yöntemi ile analiz edilmesidir. Logaritmik doğrusal modeller ile çeşitli düzeylere sahip kategorik değişkenlerin sadece temel (ana) etkileri değil, aynı zamanda değişkenler arasındaki ikili veya daha yüksek dereceden etkileşim etkileri de sınanmaktadır.

3.1.2. Kapsam

Araştırmanın örneklemini Ankara ilinde çeşitli lise ve üniversitelerde okuyan 15-25 yaş arasındaki 2343 öğrenci oluşturmaktadır. Bu kapsamda, yazarlarından izin alınarak, "Lise ve Üniversite Öğrencilerinde İntihar Riskini Belirlemeye Yönelik Bir Modelin Sınanması" (Hisli Şahin & Durak Batıgün, 2009) çalışmasının verileri kullanılmıştır.

3.1.3. Yöntem

Çalışma kapsamında ilgili değişkenler arasındaki temel etki, ikili ve daha yüksek dereceden etkileşim etkileri Logaritmik Doğrusal Modeller (Logaritmik Doğrusal Analiz) yöntemi ile "IBM SPSS Statistics 21" paket programı kullanılarak analiz edilmiştir. Ayrıca analiz yöntemine geçilmeden önce ilgili değişkenlere ait örneklem büyüklüklerine, değişkenlere ilişkin betimsel istatistiklere ve bazı klasik istatistiksel analiz yöntemlerine (t-Testi, Tek Yönlü ve İki Yönlü Varyans Analizi) yer verilmiştir.

3.2. VERİ TOPLAMA ARAÇLARI

Bu çalışmada, Hisli Şahin & Durak Batıgün (2009)'ün çalışmasındaki bireylerin intihar olasılığı ve problem çözme becerilerini belirlemek amacıyla kullanılan İntihar Olasılığı Ölçeği (İOÖ) ve Problem Çözme Envanteri (PÇE) Ölçeği verilerinden yararlanılmıştır. Bunların yanı sıra katılımcılar hakkında bazı demografik bilgilere ulaşmak amacıyla araştırmacılar tarafından hazırlanan ve uygulanan "Demografik Bilgi Formu"ndan elde edilen veriler kullanılmıştır.

3.2.1. Demografik Bilgi Formu

Bireylere ilişkin yaş, cinsiyet, öğrenim düzeyi, anne eğitim durumu, baba eğitim durumu vb. türden demografik özellikleri belirlemek amacıyla araştırmacılar tarafından hazırlanmış anket formudur.

3.2.2. İntihar Olasılığı Ölçeği

İntihar Olasılığı Ölçeği (İOÖ), 1-4 arasında puanlanan Likert tipi bir ölçek türü olup Cull ve Gill (1988) tarafından geliştirilmiştir. 36 maddeden oluşan bu ölçek bireylerin kendilerini değerlendirmelerine yönelik bir ölçek türüdür. 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)'ün yapmış olduğu bir çalışmada kullanılan formdur.

3.2.3. Problem Çözme Envanteri

Problem Çözme Envanteri (PÇE) 35 maddelik 1-6 arası Likert tipi kendini değerlendirme türü bir ölçek olup, Heppner ve Petersen (1982) tarafından geliştirilmiştir. Söz konusu ölçek bireylerin problem çözme yetileri konusunda kendi algılayışını ölçer ve ölçekten alınan yüksek puan bireylerin problem çözme yetileri konusunda kendilerini yetersiz olarak algıladıklarını yani bireylerin problem çözme becerilerinin düşük olduğunu ifade eder. Ölçeğin kültürümüze uyarlanması Şahin ve arkadaşları (1993) tarafından yapılmıştır.

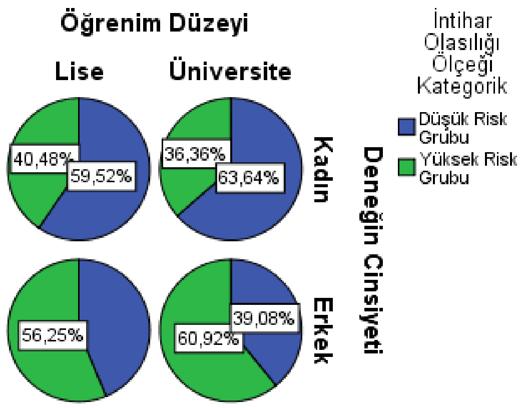
3.3. ÖZET İSTATİSTİKLER VE İSTATİSTİKSEL ANALİZ SONUÇLARI

Araştırmaya katılan öğrencilerin sosyo-demografik özelliklerinin frekans dağılımları aşağıdaki tabloda özetlenmiştir.

Tablo 3.1. Araştırmaya Katılan Öğrencilerin Sosyo-Demografik Özelliklerinin Frekans Dağılımları

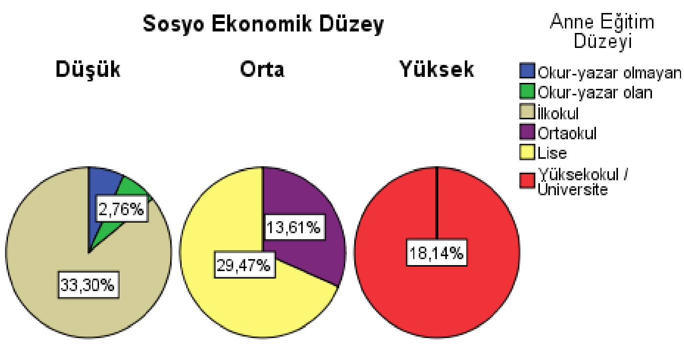
Değişken ve Düzeyleri		**Frekans**	**Yüzde**
Cinsiyet	**Kadın**	1320	%56.3
	Erkek	995	%42.5
Toplam		**2315**	**%98.8**
Yaş	**15 – 17**	1229	%52.5
	18 – 25	1047	%44.7
Toplam		**2276**	**%97.1**
Öğrenim Düzeyi	**Lise**	1358	%58.0
	Üniversite	985	%42.0
Toplam		**2343**	**%100**
Anne Eğitim Düzeyi	**Okur-yazar 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
Toplam		**2321**	**%99.1**
Sosyo Ekonomik Düzey	**Düşük**	900	%38.4
	Orta	1000	%42.6
	Yüksek	421	%19.0
Toplam		2321	%99.1
Annenin sağ olup olmama durumu	**Sağ**	2318	%98.9
	Hayatta değil	18	%0.8
Toplam		**2336**	**%99.7**
Babanın sağ olup olmama durumu	**Sağ**	2250	%96.0
	Hayatta değil	85	%3.6
Toplam		**2335**	**%99.7**
Aile Birliği	**Birlikte**	2106	%89.9
	Ayrı	134	%5.7
Toplam		**2240**	**%95.6**

Araştırmanın örneklemini Ankara ilinde çeşitli lise ve üniversitelerde okuyan 2343 öğrenci oluşturmaktadır. Bunların 1358'i lise (%58.0), 985'i üniversite (%42.0) öğrencisidir. Örneklemin %56.3 'ü kız, %42.5'u erkektir. Bireylerin yaş aralığı 15-25 olup, 15 ile 17 yaş arasında olanlar "ergenlik" (%52.5), 18 ile 25 yaş arasında olanlar "geç ergenlik" (%44.7) olarak sınıflandırılmıştır.



Şekil 3.1. Cinsiyet × Öğrenim Düzeyi × İntihar Olasılığı Ölçeği Değişkenlerinin Dağılımı

Anne eğitim düzeyi, sosyo ekonomik 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 sadece sosyo ekonomik düzeyi düşük ve yüksek olan gruplar değerlendirilmiştir.



Şekil 3.2. Sosyo Ekonomik Düzey × Anne Eğitim Düzeyi Değişkenlerinin Dağılımı

Çalışmaya katılan bireylerin %98.9'unun annesi hayatta iken sadece %0.8'inin annesi hayatta değildir. Aynı şekilde %96.0'sının babası hayatta iken, sadece %3.6'lık kısmın babası hayatta değildir ve bireylerin %94'ü ailesiyle birlikte yaşarken, %6'sı ailesinden ayrı yaşamaktadır. Çalışmaya katılan bireylerin neredeyse tamamının en az annesi ya da babası sağ iken, büyük bir çoğunluğun da ailesiyle birlikte yaşadıklarını söyleyebiliriz.

Örneklemin intihar olasılığı ölçeği toplam puanı ortalaması, $\bar{X} = 73.07$ ve standart sapması, $ss = 12.25$ olup; intihar olasılığı puanı ortalamanın 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. Sosyo ekonomik düzeyi orta grupta yer alan bireyler çalışmaya logaritmik doğrusal analiz yönteminde dahil edilmemiştir.

Logaritmik Doğrusal Analiz yöntemine geçilmeden önce sürekli değişkenler intihar olasılığı Ölçeği Ortalama Toplam Puanı bağımlı değişken olarak alınarak bazı demografik değişkenlere (cinsiyet, aile birliği, sosyo ekonomik düzey) yönelik analizler (t-Testi, tek yönlü ve iki yönlü varyans analizi) yapılmış ve sonuçlar aşağıdaki gibi özetlenmiştir.

İntihar olasılığı ölçeği ortalama toplam puanının cinsiyete göre anlamlı farklılık gösterip göstermediğine ilişkin Bağımsız Örneklemler t-Testi sonuçları aşağıdaki tabloda verilmiştir.

Tablo 3.2. İntihar Olasılığı Ölçeği Ortalama Toplam Puanının Bağımsız Örneklemler t-Testi Sonuçları

Cinsiyet	N	Ort.	ss	sd	t	p
Kadın	1320	72.14	11.118	2313	-4.475	0.000
Erkek	995	74.19	10.633			

Elde edilen bulgulara göre intihar olasılığı ölçeği ortalama toplam puanları arasında cinsiyete göre istatistiksel olarak anlamlı bir farklılık bulunmuştur ($t_{(2313)} = -4.475; p < 0.05$). Erkeklerin intihar olasılığı ölçeği ortalama toplam puanı, ($\bar{x} = 74.19; ss = 10.63$) kadınların intihar olasılığı ölçeği ortalama toplam puanı ortalamasından ($\bar{x} = 72.14; ss = 11.12$) daha yüksek çıkmıştır. Bu sonuçlar, erkeklerin kadınlardan daha fazla ortalama toplam puanı üstü değer taşıdığını yani yüksek risk grubunda yer aldığını gösterir.

İntihar olasılığı ölçeği ortalama toplam puanının aile birliğine göre anlamlı farklılık gösterip göstermediğine ilişkin Bağımsız Örneklemler t-Testi sonuçları aşağıdaki tabloda verilmiştir.

Tablo 3.3. İntihar Olasılığı Ölçeği Ortalama Toplam Puanının Bağımsız Örneklemler t-Testi Sonuçları

Aile Birliği	N	Ort.	ss	sd	t	p
Evet	2106	73.01	10.986	2238	-1.544	0.123
Hayır	134	74.53	11.772			

Elde edilen bulgulara göre intihar olasılığı ölçeği ortalama toplam puanları arasında istatistiksel olarak anlamlı bir farklılık bulunmamıştır ($t_{(2238)} = -1.544; p > 0.05$). Bu sonuç gruplar arasında fark olmadığını, yani anne ya da baba bile birlikte yaşayıp yaşamamanın intihar olasılığı ölçeği ortalama toplam puanı değerleri arasında istatistiksel olarak anlamlı bir farklılık taşımadığını gösterir. Bu nedenle logaritmik doğrusal analizlerde anlamlı farklılık taşımayan bu değişkenler değerlendirilmemiştir.

İntihar olasılığı ölçeği ortalama toplam puanının sosyo ekonomik düzeylere göre anlamlı bir farklılık gösterip göstermediğine ilişkin tek yönlü varyans analizi sonuçları aşağıdaki tablolarda verilmiştir.

Tablo 3.4. Tanımlayıcı İstatistikler

İntihar Olasılığı Ölçeği Toplam Puanı	N	$\bar{x}$	ss
Düşük	257	74.46	18.715
Orta	297	72.78	18.666
Yüksek	124	71.51	17.972
Toplam	678	73.94	18.642

Tablo 3.5. İntihar Olasılığı Ölçeği Ortalama Toplam Puanının Tek Faktörlü Varyans Analizi Sonuçları

	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Gruplar arası	2755.821	2	1377.910	4.000	0.019
Gruplar İçi	232520.936	675	344.475		
Toplam	235276.757	677			

Tablo 3.6. İntihar Olasılığı Ölçeği Ortalama Toplam Puanının Çoklu Karşılaştırma Tukey Testi Sonuçları

	(I) Sosyo Ekonomik Düzey	(J) Sosyo Ekonomik Düzey	Ortalama Fark (I-J)	*ss*	*p*	95% Güven Aralığı	
						Alt Sınır	Üst Sınır
Tukey HSD	Düşük	Orta	3.671	1.581	0.054	-0.04	7.38
		Yüksek	4.947*	2.029	0.040	0.18	9.71
	Orta	Düşük	-3.671	1.581	0.054	-7.38	0.04
		Yüksek	1.276	1.984	0.796	-3.38	5.94
	Yüksek	Düşük	-4.947*	2.029	0.040	-9.71	-0.18
		Orta	-1.276	1.984	0.796	-5.94	3.38

Tek Faktörlü Varyans Analizi sonuçlarına göre sosyo ekonomik düzeyi farklı olan kişilerin intihar olasılığı ölçeği ortalama toplam puanları arasında istatistiksel olarak anlamlı farklılık vardır ($F_{(2,675)} = 4.000, p < 0.05$). Farklılığın hangi ikili grup ya da gruplardan kaynaklandığını gösteren Tukey testi sonuçlarına bakıldığında sosyo ekonomik düzeyi düşük olanlarla sosyo ekonomik düzeyi yüksek olanların intihar olasılığı ölçeği ortalama toplam puanları arasında istatistiksel olarak anlamlı bir farklılık olduğu görülmüştür. Yani sosyo ekonomik düzeyi düşük olan grubun intihar olasılığı ölçeği ortalama toplam puanı ($\bar{x} = 74.46, ss = 18.715$), sosyo ekonomik düzeyi yüksek olan gruba ($\bar{x} = 71.51, ss = 17.972$) göre daha yüksektir.

İntihar olasılığı ölçeği ortalama toplam puanının cinsiyet ve sosyo ekonomik düzeylere göre anlamlı bir farklılık gösterip göstermediğine ilişkin çift yönlü varyans analizi sonuçları;

Tablo 3.7. Tanımlayıcı İstatistikler

Deneğin Cinsiyeti	Sosyo Ekonomik Düzey	$\bar{x}$	ss	Örneklem
Kadın	Düşük	73.415	12.1665	419
	Orta	71.184	12.1695	485
	Yüksek	70.476	11.5867	187
	Toplam	71.919	12.1199	1091
Erkek	Düşük	75.725	12.4578	284
	Orta	74.171	12.0205	327
	Yüksek	72.940	11.4771	151
	Toplam	74.507	12.1106	762
Total	Düşük	74.349	12.3285	703
	Orta	72.387	12.1908	812
	Yüksek	71.577	11.5859	338
	Toplam	72.983	12.1796	1853

Tablo 3.8. Cinsiyet ve Sosyo Ekonomik Düzey Değişkenlerine Ait Çift Yönlü Varyans Analizi Sonuçları

Kaynak	Tip III Ortalama Kareler	*sd*	Ortalama Kareler	*F*	*p*	Kısmi Eta Kareler
Düzeltilmiş Model	5421.979[a]	5	1084.396	7.437	0.000	0.020
Kesme Noktası	8338467.665	1	8338467.665	57188.184	0.000	0.969
Cinsiyet	2619.948	1	2619.948	17.969	0.000	0.010
Sesgroup	2228.233	2	1114.116	7.641	0.000	0.008
Cinsiyet * Sesgroup	44.603	2	22.302	0.153	0.858	0.000
Hata	269306.503	1847	145.808			
Toplam	10144840.000	1853				
Düzeltilmiş Toplam	274728.481	1852				

tablolarında özetlenmiştir. Sonuçlar incelendiğinde, cinsiyet ve sosyo ekonomik düzeyinin intihar olasılığı ölçeği ortalama toplam puanında ana (temel) etkileri anlamlı bulunurken $(F_{(1)} = 17.969; F_{(2)} = 7.641; p < 0.05)$, birlikte etkileşiminin intihar olasılığı ölçeği ortalama toplam puanı üzerindeki etkisi istatistiksel olarak anlamlı bulunmamıştır $(F_{(3)} = 0.153; p > 0.05)$. Ancak Tablo 3.8'de kısmi eta kareler değerleri incelendiğinde temel etkileri istatistiksel olarak anlamlı bulunan değişkenlerin (cinsiyet, sosyo ekonomik düzey) etkisinin çok büyük olduğu söylenemez. Çünkü bu etkilerin intihar olasılığı ölçeği ortalama toplam puanı üzerindeki etkisi sırasıyla % 1 ve % 0.8'dir.

3.4. LOGARİTMİK DOĞRUSAL ANALİZ YÖNTEMİ

Bu bölümde klasik istatistiksel analiz yöntemi sonuçları da göz önünde bulundurularak bazı değişkenler arasındaki ikili veya daha yüksek dereceden etkileşimler çalışmanın temel amacını oluşturan logaritmik doğrusal modeller yardımı ile incelenmiştir.

3.4.1. Üç Yönlü Çapraz Tabloların Logaritmik Doğrusal Analiz Yöntemi İle İncelenmesi

"cinsiyet", "sosyo ekonomik düzey" ve "intihar olasılığı ölçeği" değişkenlerine ilişkin logaritmik doğrusal analiz

Değişken düzeyleri ve kodlamaları aşağıdaki tabloda verilmiştir.

Tablo 3.9. Değişkenler ve Kategorileri

Değişken	**Düzeyleri**
Cinsiyet	Kadın = 1
	Erkek = 2
İntihar Olasılığı Ölçeği (iookateg)	Düşük Risk Grubu = 1
	Yüksek Risk Grubu = 2
Sosyo Ekonomik Düzey (sesgroup)	Düşük = 1
	Orta = 2
	Yüksek = 3

İlgili değişkenlere geriye doğru aşamalı logaritmik doğrusal analiz uygulanmış ve ana etkiler (K=1), iki değişkenli etkileşim etkileri (K=2), üç değişkenli etkileşim etkilerinin (K=3) anlamlılığına ait analiz sonucu aşağıda verilmiştir.

Tablo 3.10. 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 3.10’daki ilk kısmın birinci satırı incelendiğinde ana etkiler, ikili etkileşim etkileri ve üçlü etkileşim etkilerinin sıfır olduğunu ifade eden H_0 yokluk hipotezi,

$$H_0 = \lambda_{ijk}^{ABC} = 0, \qquad \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = 0, \qquad \lambda_i^A = \lambda_j^B = \lambda_k^C = 0$$

Gerek olabilirlik oranı ($L^2 = 65.638, p = 0.000 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 68.723, p = 0.000 < 0.05$) değerleri 0.05 anlam düzeyinde red edilmiştir.

İkinci satırda ise, ikinci dereceden ve üçüncü dereceden etkilerin sıfıra eşit olduğu H_0 yokluk hipotezi,

$$H_0 = \lambda_{ijk}^{ABC} = 0, \qquad \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = 0$$

Gerek olabilirlik oranı ($L^2 = 21.549, p = 0.003 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 21.454, p = 0.003 < 0.05$) değerleri 0.05 anlam düzeyinde red edilmiştir.

Son satırda ise, üçüncü dereceden etkilerin sıfıra eşit olduğunu ifade eden H_0 yokluk hipotezi,

$$H_0 = \lambda_{ijk}^{ABC} = 0$$

Gerek olabilirlik oranı ($L^2 = 0.055, p = 0.973 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 0.055, p = 0.003 < 0.05$) değerleri 0.05 anlam düzeyinde red edilememiştir.

Tablo 3.10'un ikinci kısmı ise, sırasıyla ana etkiler, ikinci dereceden ve üçüncü dereceden etkilerin sıfıra eşit olduğu hipotezlerini test eder. Ana etkiler ve ikinci dereceden etkileşim etkilerinin sıfıra eşit olduğu H_0 yokluk hipotezi, gerek olabilirlik oranı, gerekse Pearson Ki-Kare test istatistikleri p değerleri 0.05 anlam düzeyinde red edilir ($p = 0.000\ ve\ p = 0.001 < 0.05$). Ancak üçüncü dereceden etkileşim etkisinin sıfıra eşit olduğu H_0 yokluk hipotezi, gerek olabilirlik oranı, gerekse Pearson Ki-Kare test istatistiği p değeri, 0.05 anlam düzeyinde ($p = 0.973 > 0.05$) red edilememiştir. 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 modele benzer olacağını göstermektedir.

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (3.1)$$

Modelde yer alacak ikinci dereceden etkileşim parametrelerini tespit etmek için aşağıdaki kısmi ilişkiler test istatistikleri tablosu incelenebilir.

Tablo 3.11. Etki ve Etkileşimlere Ait Kısmi Ki-Kare ve Olasılık Değerleri

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

Tablo 3.11. incelendiğinde ana etkilerden cinsiyet, sesgroup; ikinci dereceden etkileşim etkisi parametrelerinden cinsiyet*iookateg, sesgroup*iookateg 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 gerekir. Bu nedenle kısmi ilişkiler tablosunda 0.05 anlam düzeyinde anlamlı bulunmayan (ioökateg) ana etkisi parametresinin, cinsiyet ve sosyo ekonomik düzey (sesgroup) değişkenleri arasında ikili etkileşim etkisinin olması nedeni ile modele dâhil edilmesi gerektiği tespit edilmiştir. Bu durumda verilere en uygun model;

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (3.2)$$

şeklinde ifade edilen koşullu bağımsızlık modeli olacaktır.

Ana etkiler ve etkileşimlere ait parametre tahminleri aşağıdaki tabloda özetlenmiştir.

Tablo 3.12. Etki ve Etkileşimlere Ait Parametre Tahminleri

Etki	Parametre	Tahmin	Std. Hata	*Z*	*p*	%95 Güven Aralığı	
						Alt Sınır	Üst Sınır
cinsiyet*iookateg*sesgroup	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*iookateg	1	0.207	0.071	2.933	0.003	0.069	0.345
cinsiyet*sesgroup	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
iookateg*sesgroup	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
iookateg	1	0.061	0.071	0.864	0.388	-0.077	0.199
sesgroup	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

Tablo 3.12 incelendiğinde uygun modelde yer alan bazı değişkenlerin (cinsiyet, sosyo ekonomik düzey, cinsiyet ve intihar olasılığı ölçeği etkileşim etkisi) p değerlerinin 0.05’den küçük olduğu görülmektedir.

Ayrıca standartlaştırılmış parametre tahminleri göz önünde bulundurulup Tablo 3.12 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; sosyo ekonomik düzeyi yüksek olanların cinsiyet değişkeninin yine kadın kategorisinde bağımlı olduğu söylenebilir.

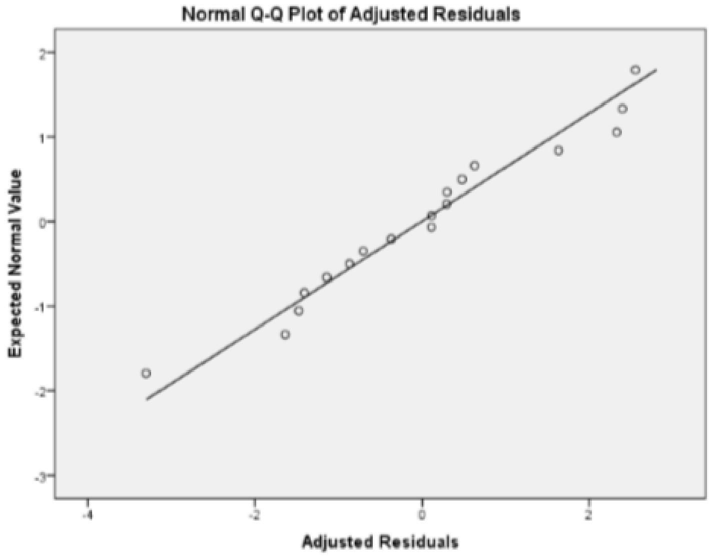
Gözlenen frekanslar ile beklenen frekanslar arasındaki fark yani model artıkları seçilen modele ilişkin çıkarımlar yapabilmemizi sağlayan bir ölçüttür. Cinsiyet, Sosyo Ekonomik Düzey ve İntihar Olasılığı Ölçeği değişkenleri için model artıkları aşağıdaki tabloda verilmiştir.

Tablo 3.13. Cinsiyet, İntihar Olasılığı Ölçeği ve Sosyo Ekonomik Düzey Değişkenlerine İlişkin Model Artıkları

Cinsiyet	İntihar Olasılığı Ölçeği Kategorik	Sosyo Ekonomik Düzey	Gözlenen		Beklenen		Artıklar	Std. Artıklar
			Frekans	%	Frekans	%		
Kadın	Düşük Risk Grubu	Düşük	32.000	12.6%	33.075	13.1 %	-1.075	-0.187
		Orta	46.000	18.2%	47.775	18.9 %	-1.775	-0.257
		Yüksek	20.000	7.9%	17.150	6.8%	2.850	0.688
	Yüksek Risk Grubu	Düşük	22.000	8.7%	20.925	8.3%	1.075	0.235
		Orta	32.000	12.6%	30.225	11.9 %	1.775	0.323
		Yüksek	8.000	3.2%	10.850	4.3%	-2.850	-0.865
Erkek	Düşük Risk Grubu	Düşük	16.000	6.3%	17.978	7.1%	-1.978	-0.467
		Orta	11.000	4.3%	11.441	4.5%	-0.441	-0.130
		Yüksek	11.000	4.3%	8.581	3.4%	2.419	0.826
	Yüksek Risk Grubu	Düşük	28.000	11.1%	26.022	10.3 %	1.978	0.388
		Orta	17.000	6.7%	16.559	6.5%	0.441	0.108
		Yüksek	10.000	4.0%	12.419	4.9%	-2.419	-0.687

Tablo 3.13 incelendiğinde artıkların değerlerinin küçük olduğu ve standartlaştırılmış artıkların mutlak değerce ikiyi geçmediği görülmektedir. Bu sonuç elde edilen modelin verilere uygun bir model olduğunu göstermektedir.

Standartlaştırılmış artıklara ilişkin normal olasılık grafiği ise aşağıdaki şekilde verilmiştir.



Şekil 3.3. Standartlaştırılmış Artıklara İlişkin Normal Olasılık Grafiği

Şekil 3.3. incelendiğinde de standartlaştırılmış artıkların normal dağılımdan sapmadığı ve elde edilen modelin verilere en uygun model olduğu tespit edilmiştir.

Cinsiyet ve İntihar Olasılığı Ölçeği Değişkenlerinin Odds Oranı İle Yorumlanması

Cinsiyet ve intihar olasılığı ölçeği değişkenlerine ait parametre tahminleri aşağıdaki tabloda verilmiştir.

Tablo 3.14. Cinsiyet ve İntihar Olasılığı Ölçeği Parametre Tahminleri

Parametre	**Tahmin**	**Std.Hata**	***Z***	***p***	**%95 Güven Aralığı**	
					Alt Sınır	**Üst Sınır**
Sabit	5.085^{a}					
[cinsiyet=1]	-0.044	0.113	-0.394	0.694	-0.265	0.176
[cinsiyet=2]	0^{b}	.	.	.	.	.
[iookateg=1]	-0.310	0.121	-2.559	0.010	-0.547	-0.073
[iookateg=2]	0^{b}	.	.	.	.	.
[cinsiyet=1]*[iookateg=1]	0.756	0.159	4.760	0.000	0.445	1.068
[cinsiyet=1]*[iookateg=2]	0^{b}	.	.	.	.	.
[cinsiyet=2]*[iookateg=1]	0^{b}	.	.	.	.	.
[cinsiyet=2]*[iookateg=2]	0^{b}	.	.	.	.	.

Tablo 3.14 incelendiğinde [cinsiyet=1]*[iookateg=1] için hesaplanan λ katsayısı 0.756 bulunmuştur. 0.756 değeri, anti-logaritması alındığında $\{e^{0.756} = 2.129 \cong 2.13\}$, örnek odds oranı $\hat{\theta}$'ya eşittir.

Örnek odds oranı $\hat{\theta}$, parametre tahminlerinden hesaplanabileceği gibi çapraz tablo yardımıyla da hesaplanabilir.

Tablo 3.15. Cinsiyet ve İntihar Olasılığı Ölçeği Değişkenlerine Ait Çapraz Tablo

		İntihar Olasılığı Ölçeği Kategorik		Toplam
		Düşük Risk Grubu	**Yüksek Risk Grubu**	
Deneğin Cinsiyeti	**Kadın**	241	154	395
	Erkek	118	161	279
Toplam		359	315	674

Örnek odds oranı,

$$\hat{\theta} = \frac{odds_1}{odds_2} = \frac{Kadın\ (Düşük/Yüksek)}{Erkek\ (Düşük/Yüksek)} = \frac{241/154}{118/161} = 2.135$$

şeklinde hesaplanır. Bu oran cinsiyeti kadın olanların düşük riskli intihar grubunda olma oranı, erkeklerin düşük riskli intihar grubunda olma oranından 2.13 kat daha fazladır şeklinde yorumlanır.

Odds oranı için %95 güven aralığı;

$\text{Log}(2.135) \mp 1.96\sqrt{\frac{1}{241}+\frac{1}{154}+\frac{1}{118}+\frac{1}{161}}$ şeklinde hesaplanır ve

$(0.7585) \mp 1.96(0.1591)$ ya da $(0.4467, 1.0703)$ şeklinde tanımlanır. Bu aralık limitlerin üssü alınarak $(e^{0.4467}, e^{1.0703})$ hesaplandığında $(1.56, 2.92)$ olarak bulunur. Güven aralığı değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içermediğinden hesaplanan odds oranının istatistiksel olarak anlamlı olduğu dolayısıyla gruplar arasında farklılık olduğu tespit edilmiştir.

Sosyo Ekonomik Düzey ve İntihar Olasılığı Ölçeği Değişkenlerinin Odds Oranı İle Yorumlanması

Sosyo ekonomik düzey ve intihar olasılığı ölçeği değişkenlerine ait parametre tahminleri aşağıdaki tabloda verilmiştir.

Tablo 3.16. Sosyo Ekonomik Düzey ve İntihar Olasılığı Ölçeği Parametre Tahminleri

Parametre	Tahmin	Std.Hata	*Z*	*p*	%95 Güven Aralığı	
					Alt Sınır	Üst Sınır
Sabit	3.942^{a}					
[iookateg=1]	0.356	0.182	1.957	0.050	0.000	0.712
[iookateg=2]	0^{b}	.	.	.	.	.
[sesgroup=1]	0.982	0.163	6.011	0.000	0.662	1.302
[sesgroup=2]	0.930	0.165	5.650	0.000	0.607	1.252
[sesgroup=3]	0^{b}	.	.	.	.	.
[iookateg=1]*[sesgroup=1]	−0.488	0.220	−2.212	0.027	−0.920	−0.056
[iookateg=1]*[sesgroup=2]	−0.106	0.216	−0.491	0.623	−0.529	0.317
[iookateg=1]*[sesgroup=3]	0^{b}	.	.	.	.	.
[iookateg=2]*[sesgroup=1]	0^{b}	.	.	.	.	.
[iookateg=2]*[sesgroup=2]	0^{b}					
[iookateg=2]*[sesgroup=3]	0^{b}	.	.	.	.	.

Tablo 3.16 incelendiğinde [iookateg=1]*[sesgroup=1] için hesaplanan λ katsayısı -0.488 bulunmuştur. -0.488 değeri, anti-logaritması alındığında $\{e^{-0.488} = 0.613 \cong 0.611\}$, örnek odds oranı $\hat{\theta}$'ya eşittir.

Örnek odds oranı $\hat{\theta}$, parametre tahminlerinden hesaplanabileceği gibi çapraz tablo yardımıyla da hesaplanabilir.

Tablo 3.17. Sosyo Ekonomik Düzey ve İntihar Olasılığı Ölçeği Değişkenlerine Ait Çapraz Tablo

		İntihar Olasılığı Ölçeği Kategorik		Toplam
		Düşük	**Yüksek**	
Sosyo Ekonomik Düzey	**Düşük**	120	137	257
	Yüksek	73	51	124
Toplam		193	188	381

Örnek odds oranı,

$$\hat{\theta} = \frac{odds_1}{odds_2} = \frac{Düşük\ (Düşük/Yüksek)}{Yüksek\ (Düşük/Yüksek)} = \frac{120/137}{73/51} = 0.611$$

şeklinde hesaplanır. Bu oran, $1/0.61 = 1.63$ olarak alındığında, sosyo ekonomik düzeyi yüksek olan grubun düşük riskli intihar grubunda olma oranı, sosyo ekonomik düzeyi düşük olan grubun düşük riskli intihar grubunda olma oranından 1.63 kat daha fazladır şeklinde yorumlanır.

Odds oranı için %95 güven aralığı;

$\text{Log}(0.611) \mp 1.96\sqrt{\frac{1}{120}+\frac{1}{137}+\frac{1}{73}+\frac{1}{51}}$ şeklinde hesaplanır ve

$(-0.4927) \mp 1.96(0.2211)$ ya da $(-0.9261, -0.0593)$ şeklinde tanımlanır. Bu aralık limitlerin üssü alınarak $(e^{-0.9261}, e^{0.0593})$ hesaplandığında (0.40,1.06) olarak bulunur. Güven aralığı, değişkenlerin birbirinden bağımsız olduğunu ifaden eden $\theta = 1$ değerini içerdiğinden hesaplanan odds oranının istatistiksel olarak anlamlı olmadığı ve gruplar arasında farklılık olmadığı tespit edilmiştir.

Burada dikkat edilmesi gereken iki yönlü log-lineer modellerde parametre tahmin değerlerinin anti-logaritmasının, iki yönlü çapraz tablolardan hesaplanan, örnek odds oranı $\hat{\theta}$'ya eşitliğidir.

“cinsiyet”, “öğrenim düzeyi” ve “problem çözme envanteri” değişkenlerine ilişkin logaritmik doğrusal analiz

“Problem Çözme Envanteri (PÇE)” ölçekten alınan toplam puanlara göre düşük, yüksek ve orta olarak gruplandırılmıştır. Ancak PÇE ölçeği toplam puanı orta grupta yer alanlar analizlerde değerlendirilmemiştir. PÇE ölçeğinden alınan yüksek puan, bireylerin problem çözme yetileri konusunda kendilerini yetersiz olarak algıladıklarını yani bireylerin problem çözme becerilerinin düşük olduğunu ifade eder.

Değişken düzeyleri ve kodlamaları aşağıdaki tabloda verilmiştir.

Tablo 3.18. Değişkenler ve Kategorileri

Değişken	**Düzeyleri**
Cinsiyet	Kadın = 1
	Erkek = 2
Öğrenim Düzeyi (tür)	Lise = 1
	Üniversite = 2
Problem Çözme Envanteri (pcekateg)	Düşük = 1
	Yüksek= 2
	Orta=3

İlgili değişkenlere geriye doğru aşamalı logaritmik doğrusal analiz uygulanmış ve ana etkiler (K=1), iki değişkenli etkileşim etkileri (K=2), üç değişkenli etkileşim etkilerinin (K=3) anlamlılığına ait analiz sonucu aşağıda verilmiştir.

Tablo 3.19. 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	7	36.164	0.000	36.845	0.000	0
	2	4	15.704	0.003	15.626	0.004	2
	3	1	13.863	0.000	13.831	0.000	2
K-Yönlü Etkileşimler	1	3	20.460	0.000	21.219	0.000	0
	2	3	1.841	0.606	1.795	0.616	0
	3	1	13.863	0.000	13.831	0.000	0

Tablo 3.19 incelendiğinde ilk kısmın birinci satırı ana etkiler, iki yönlü etkileşim etkileri ve üç yönlü etkileşim etkilerinin sıfıra eşit olduğunu ifade eden H_0 yokluk hipotezi,

$H_0 = \lambda_{ijk}^{ABC} = 0, \qquad \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = 0, \qquad \lambda_i^A = \lambda_j^B = \lambda_k^C = 0$

Gerek olabilirlik oranı ($L^2 = 36.164, p = 0.000 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 36.845, p = 0.000 < 0.05$) p değerinin 0.000 olması sonucu red edilmektedir ($p = 0.000 < 0.05$).

İkinci satırda ise, bütün ikinci dereceden ve üçüncü dereceden etkilerin sıfıra eşit olduğunu ifade eden H_0 yokluk hipotezi,

$H_0 = \lambda_{ijk}^{ABC} = 0, \qquad \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = 0$

Gerek olabilirlik oranı ($L^2 = 15.704, p = 0.003 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 15.626, p = 0.004 < 0.05$) p değerinin 0.000 olması sonucu red edilmektedir ($p = 0.000 < 0.05$).

Son satırda ise, üçüncü dereceden etkilerin sıfıra eşit olduğunu ifade eden H_0 yokluk hipotezi,

$H_0 = \lambda_{ijk}^{ABC} = 0$

Gerek olabilirlik oranı ($L^2 = 13.863, p = 0.000 < 0.05$), gerekse Pearson Ki-Kare test istatistiği ($\chi^2 = 13.801, p = 0.000 < 0.05$) p değerinin 0.000 olması sonucu red edilmektedir ($p = 0.000 < 0.05$).

Tablo 3.19'un ikinci kısmı ise, sırasıyla ana etkiler, ikinci dereceden ve üçüncü dereceden etkilerin sıfıra eşit olduğu hipotezlerini test eder. Bu hipotezler de gerek Gerek olabilirlik oranı, gerekse Pearson Ki-Kare test istatistikleri p değerinin 0.000 olması sonucu red edilir ($p = 0.000 < 0.05$). Bu sonuçlar verilere en uygun modelin ana etkileri, tüm ikili ve üçlü değişkenlerin birlikte etkileşimlerini içeren Eşitlik 3.2.'de verilen doymuş hiyerarşik logaritmik doğrusal model olacağını gösterir.

$$\log(m_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \qquad (3.2)$$

Modelde yer alacak ikinci dereceden etkileşim parametrelerini test edebilmek için Tablo 3.20'de verilen kısmi ilişkiler test istatistikleri tablosu incelebilir.

Tablo 3.20. Etki ve Etkileşimlere Ait Kısmi Ki-Kare ve Olasılık Değerleri

Etki	sd	Kısmi Ki-Kare	p	İterasyon Sayısı
cinsiyet*tür	1	0.086	0.770	2
cinsiyet*pcekateg	1	1.066	0.302	2
tür*pcekateg	1	0.651	0.420	2
cinsiyet	1	17.734	0.000	2
tür	1	2.326	0.127	2
pcekateg	1	0.400	0.527	2

Tablo 3.20 incelendiğinde ana etkilerden sadece cinsiyet değişkeni 0.05 anlam düzeyinde istatistiksel olarak anlamlı bulunurken $(p < 0.05)$, diğer değişkenler ve ikili etkileşim etkileri istatistiksel olarak anlamlı bulunmamıştır $(p > 0.05)$. Ancak hiyerarşik logaritmik doğrusal modellere göre yüksek dereceli bir terim modelde var ise, daha düşük tüm terimlerin de modelde yer alması gerekir.

Değişkenlere ilişkin ana etkiler ve etkileşimlere ait parametre tahminleri aşağıdaki gibidir.

Tablo 3.21. Etki ve Etkileşimlere Ait Parametre Tahminleri

Etki	Parametre	Tahmin	Std. Hata	Z	p	95% Güven Aralığı	
						Alt Sınır	Üst Sınır
cinsiyet*pcekateg*tür	1	0.140	0.038	3.696	0.000	0.066	0.215
cinsiyet*pcekateg	1	0.032	0.038	0.838	0.402	-0.043	0.106
cinsiyet*tür	1	0.016	0.038	0.432	0.666	-0.058	0.091
pcekateg*tür	1	0.008	0.038	0.218	0.828	-0.066	0.083
cinsiyet	1	0.156	0.038	4.102	0.000	0.081	0.230
pcekateg	1	-0.032	0.038	-0.847	0.397	-0.107	0.042
tür	1	0.050	0.038	1.326	0.185	-0.024	0.125

Tablo 3.21 incelendiğinde ana etkilerden sadece cinsiyet değişkeni ve üçlü etkileşim etkileri 0.05 anlam düzeyinde istatistiksel olarak anlamlı bulunmuştur ($p < 0.05$). Standartlaştırılmış parametre tahminleri incelendiğinde ise en büyük Z değeri 4.102 olup cinsiyet değişkenine aittir. Bu değeri, 3.696 ile üçlü etkileşim etkileri parametresi takip eder. Yani hücre frekansına en önemli katkıyı sağlayan faktörlerin 4.102 ile cinsiyet ve 3.696 ile üçlü etkileşim etkisi (cinsiyet*pcekateg*tür) parametresine ait olduğunu görülmektedir.

Öğrenim Düzeyi ve Problem Çözme Envanteri Değişkenlerinin Odds Oranı İle Yorumlanması

Öğrenim düzeyi ve problem çözme envanteri değişkenlerine ait parametre tahminleri aşağıdaki tabloda verilmiştir.

Tablo 3.22. Öğrenim Düzeyi ve Problem Çözme Envanteri Parametre Tahminleri

Parametre	**Tahmin**	**Std. Hata**	***Z***	***p***	**%95 Güven Aralığı**	
					Alt Sınır	**Üst Sınır**
Sabit	6.467^a					
[tür=1]	0.411	0.051	8.082	0.000	0.311	0.511
[tür=2]	0^b	.	.	.	.	.
[pcekateg=1]	−1.382	0.088	−15.708	0.000	−1.555	−1.210
[pcekateg=2]	−1.266	0.084	−15.059	0.000	−1.430	−1.101
[pcekateg=3]	0^b	.	.	.	.	.
[tür=1]*[pcekateg=1]	0.112	0.118	−1.949	0.051	−0.462	0.001
[tür=1]*[pcekateg=2]	−0.337	0.115	−2.929	0.003	−0.562	−0.111
[tür=1]*[pcekateg=3]	0^b	.	.	.	.	.
[tür=2]*[pcekateg=1]	0^b	.	.	.	.	.
[tür=2]*[pcekateg=2]	0^b					
[tür=2]*[pcekateg=3]	0^b	.	.	.	.	.

Tablo 3.22 incelendiğinde [tür=1]*[pcekateg=1] için hesaplanan λ katsayısı 0.112 bulunmuştur. 0.112 değeri, anti-logaritması alındığında $\{e^{0.112} = 1.118 \cong 1.11\}$, örnek odds oranı $\hat{\theta}$'ya eşittir.

Örnek odds oranı $\hat{\theta}$, parametre tahminlerinden hesaplanabileceği gibi çapraz tablo yardımıyla da hesaplanabilir.

Tablo 3.23. Öğrenim Düzeyi ve Problem Çözme Envanteri Değişkenlerine Ait Çapraz Tablo

		Problem Çözme Envanteri Kategorik		**Toplam**
		Düşük	**Yüksek**	
Öğrenim Düzeyi	**Lise**	193	195	388
	Üniversite	161	181	342
Toplam		354	376	730

Örnek odds oranı,

$$\hat{\theta} = \frac{odds_1}{odds_2} = \frac{Lise(Düşük/Yüksek)}{Üniversite\ (Düşük/Yüksek)} = \frac{193/195}{161/181} = 1.11$$

şeklinde hesaplanır ve öğrenim düzeyi lise olan grubun problem çözme envanteri ölçeği toplam puanının düşük olma oranı, öğrenim düzeyi üniversite olan grubun problem çözme envanteri ölçeği toplam puanının düşük olma oranından 1.11 kat daha fazladır şeklinde yorumlanır. Yani lise öğrencilerinin problem çözme becerilerinin üniversite öğrencilerine göre daha yüksek olduğu tespit edilmiştir. Ayrıca Clum ve ark. (1979), stres altında bulunan bireyler katı bir bilişsel yapıya sahip ve/veya problem çözme becerilerinde bir yetersizlikleri var ise bu kişilerin intihar etme ya da intihar girişiminde bulunma olasılıklarını artırdığını ifade etmektedir. Buradan hareketle üniversite öğrencilerinin yüksek riskli intihar grubunda olması olasılığının, lise öğrencilerine göre daha yüksek olduğu söylenilebilir.

Odds oranı için %95 güven aralığı;

$\mathrm{Log}(1.11) \mp 1.96\sqrt{\frac{1}{193} + \frac{1}{195} + \frac{1}{161} + \frac{1}{181}}$ şeklinde hesaplanır ve

$0.1044 \mp 1.96(0.1483)$ ya da $(-0.1863, 0.3951)$ şeklinde tanımlanır.

Bu aralık limitlerin üssü alınarak $(e^{-0.1863}, e^{0.3951})$ hesaplandığında (0.83,1.48) olarak bulunur. Güven aralığı, değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içerdiğinden hesaplanan odds oranının istatistiksel olarak anlamlı olmadığı, dolayısıyla gruplar arasında farklılık olmadığı ifade edilir.

Öğrenim Düzeyi ve İntihar Olasılığı Ölçeği Değişkenlerinin Odds Oranı İle Yorumlanması

Öğrenim düzeyi ve intihar olasılığı ölçeği değişkenlerine ilişkin parametre tahminleri aşağıdaki tabloda verilmiştir.

Tablo 3.24. Öğrenim Düzeyi ve İntihar Olasılığı Ölçeği Parametre Tahminleri

Parametre	Tahmin	Std.Hata	Z	p	%95 Güven Aralığı	
					Alt Sınır	Üst Sınır
Sabit	4.687^{a}					
[tür=1]	0.677	0.118	5.742	0.000	0.446	0.908
[tür=2]	0^{b}	.	.	.	.	.
[iookateg=1]	0.153	0.131	1.173	0.241	-0.103	0.410
[iookateg=2]	0^{b}	.	.	.	.	.
[tür=1]*[iookateg=1]	-0.051	0.161	-0.317	0.751	-0.367	0.265
[tür=1]*[iookateg=2]	0^{b}	.	.	.	.	.
[tür=2]*[iookateg=1]	0^{b}	.	.	.	.	.
[tür=2]*[iookateg=2]	0^{b}	.	.	.	.	.

Tablo 3.24 incelendiğinde [tür=1]*[iookateg=1] için hesaplanan λ katsayısı -0.051 bulunmuştur. -0.051 değeri, anti-logaritması alındığında $\{e^{-0.051} = 0.950 \cong 0.95\}$, örnek odds oranı $\hat{\theta}$'ya eşittir.

Örnek odds oranı $\hat{\theta}$, parametre tahminlerinden hesaplanabileceği gibi çapraz tablo yardımıyla da hesaplanabilir.

Tablo 3.25. Öğrenim Düzeyi ve İntihar Olasılığı Ölçeği Değişkenlerine İlişkin Çapraz Tablo

		İntihar Olasılığı Ölçeği Kategorik		Total
		Düşük Risk Grubu	Yüksek Risk Grubu	
Öğrenim Düzeyi	**Lise**	236	213	449
	Üniversite	126	108	234
Total		362	321	683

Örnek Odds oranı,

$$\hat{\theta} = \frac{odds_1}{odds_2} = \frac{Lise\ (Düşük/Yüksek)}{Üniversite\ (Düşük/Yüksek)} = \frac{236/213}{126/108} = 0.950$$

şeklinde hesaplanır ($\theta < 1$). Bu oran $1/0.950 = 1.05$ olarak alındığında, üniversite öğrencilerinin düşük riskli intihar gurubunda olma oranı, lise öğrencilerinin düşük riskli intihar grubunda olma oranının 1.05 katıdır şeklinde yorumlanabilir. Burada intihar olasılığı düşük riskli grupta yer alan bireylerin cinsiyet değişkenine bağımlı olduğu sonucu unutulmamalıdır.

Odds oranı için %95 güven aralığı;

Log(0.950)∓(1.96)$\sqrt{\frac{1}{236} + \frac{1}{213} + \frac{1}{126} + \frac{1}{108}}$ şeklinde hesaplanır ve

$-0.0512 \mp 1.96(0.1616)$ ya da $(-0.3679, 0.2655)$ şeklinde tanımlanır.

Bu aralık limitlerin üssü alınarak $(e^{-0.3679}, e^{0.2655})$ hesaplandığında $(0.69, 1.30)$ olarak bulunur. Güven aralığı, değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içerdiğinden hesaplanan odds oranının istatistiksel olarak anlamlı olmadığı, dolayısıyla gruplar arasında farklılık olmadığı ifade edilir.

Burada dikkat edilmesi gereken iki yönlü log-lineer modellerde parametre tahmin değerlerinin anti-logaritmasının, iki yönlü çapraz tablolardan hesaplanan, örnek odds oranı $\hat{\theta}$'ya eşitliğidir.

Cinsiyet, Problem Çözme Envanteri ve Öğrenim Düzeyi Değişkenlerinin Odds Oranı İle Yorumlanması

Tablo 3.26. Cinsiyet, Öğrenim Düzeyi, Problem Çözme Envanteri Değişkenlerine İlişkin Çapraz Tablo

			Problem Çözme Envanteri	
Cinsiyet			**Düşük (1)**	**Yüksek (2)**
Kadın (1)	**Öğrenim Düzeyi**	**Lise (1)**	128	11
		Üniversite (2)	83	112
Erkek (2)	**Öğrenim Düzeyi**	**Lise (1)**	64	95
		Üniversite (2)	78	44

Koşullu odds oranları, üçüncü değişkenin sabit seviyeleri için iki değişken arasındaki oranlardır ve üçüncü değişkenin verilen değerinde diğer iki değişkenin koşullu bağımsızlığını test edebilir. Cinsiyet değişkeni için "Kadın" kategorisi sabit alındığında, öğrenim düzeyi ve problem çözme envanteri değişkeni arasındaki odds oranı,

$$\hat{\theta}_{Kadın} = \frac{odds_1}{odds_2} = \frac{Lise\ (Düşük/Yüksek)}{Üniversite\ (Düşük/Yüksek)} = \frac{128/11}{83/112} = 15.70$$

ve "Erkek" kategorisi sabit alındığında öğrenim düzeyi ve problem çözme envanteri değişkeni arasındaki odds oranı,

$$\hat{\theta}_{Erkek} = \frac{odds_1}{odds_2} = \frac{Lise\ (Düşük/Yüksek)}{Üniversite\ (Düşük/Yüksek)} = \frac{64/95}{78/44} = 0.38$$

olarak hesaplanır. Elde edilen sonuçlara göre cinsiyetin kadın olduğu biliniyor iken, lise öğrencilerinin problem çözme envanteri ölçeği toplam puanının düşük olması, üniversite öğrencilerine göre 15.70 kat daha yüksektir. Bu sonuç, cinsiyeti kadın olan lise öğrencilerinin problem çözme becerilerinin, cinsiyeti kadın olan üniversite öğrencilerine göre daha yüksek olduğunu ifade eder. Cinsiyetin erkek olduğu bilindiğinse ise; üniversite öğrencilerinin problem çözme envanteri ölçeği toplam puanının düşük olması, lise öğrencilerine göre 2.63 (=1/0.38) kat daha yüksektir. Bu sonuç, kadınların aksine cinsiyeti erkek olan üniversite öğrencilerinin problem çözme becerileri, cinsiyeti erkek olan lise öğrencilerine göre daha yüksektir.

Öğrenim düzeyi değişkeni için ‘Lise’ ve ‘Üniversite’ kategorileri sabit alındığında, cinsiyet ve problem çözme envanteri değişkeni arasındaki odds oranları da sırasıyla;

$$\hat{\theta}_{Lise} = \frac{odds_1}{odds_2} = \frac{Kadın\ (Düşük/Yüksek)}{Erkek\ (Düşük/Yüksek)} = \frac{128/11}{64/95} = 17.27$$

$$\hat{\theta}_{Üniversite} = \frac{odds_1}{odds_2} = \frac{Kadın\ (Düşük/Yüksek)}{Erkek\ (Düşük/Yüksek)} = \frac{83/112}{78/44} = 0.42$$

olarak hesaplanır. Elde edilen sonuçlara göre öğrenim düzeyinin lise olduğu biliniyor iken, kadınların problem çözme envanteri ölçeği toplam puanının düşük olması erkeklere göre 17.27 kat daha fazladır. Yani öğrenim düzeyi lise olanlarda kadınların problem çözme becerileri erkeklere göre daha yüksektir. Öğrenim düzeyinin üniversite olduğu bilindiğinde ise, erkeklerin problem çözme envanteri ölçeği toplam puanının düşük olması kadınlardan 2.38 (=1/0.42) kat daha yüksektir. Yani öğrenim düzeyi üniversite olanların, öğrenim düzeyi lise olanların aksine, erkeklerin problem çözme becerileri kadınlardan daha yüksektir.

SONUÇ ve DEĞERLENDİRME

Kategorik verilerde istatistiksel yöntemlerin kullanımı sosyal bilimlere ilişkin uygulamalarda oldukça yaygındır. Kategorik değişkenler arasındaki ikili ilişkilerin test edilmesinde kullanılan yöntemlerin başında χ^2 analizi gelmektedir. Ancak χ^2 analizi değişken sayısının ikiden fazla olması durumunda ilişki yapılarının belirlenmesinde yetersiz kalmaktadır. Bu durumda ikiden fazla değişken içeren ve değişken sayısına göre adlandırılan çapraz tablolarda ilişki yapılarının irdelenmesinde logaritmik doğrusal modellerden faydalanılmaktadır.

Bu çalışmada kategorik verilerin analizinde kullanılan ve değişkenler arasında 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 üç yönlü çapraz tablolara uygulanmıştır. Ayrıca analiz yöntemine geçilmeden önce bazı klasik istatistiksel yöntemler (t-Testi, Varyans Analizi) kullanılarak değişkenlere ilişkin bazı çıkarmalar yapılmıştır.

Çalışmada klasik istatistiksel analiz yöntemi sonuçlarından elde edilen bulgulara göre; intihar olasılığı ölçeği ortalama toplam puanının cinsiyete göre anlamlı bir farklılık yarattığı tespit edilmiştir. Ayrıca sosyo ekonomik düzeyi düşük olan bireylerin intihar olasılığı ölçeği ortalama toplam puanının, sosyo ekonomik düzeyi yüksek olan bireylere göre daha yüksek olduğu sonucuna ulaşılmıştır. Ancak intihar olasılığı ölçeği ortalama toplam puanının aile birliğine göre istatistiksel olarak anlamlı bir farklılık yaratmadığı tespit edilmiştir. Cinsiyet ve sosyo ekonomik düzey değişkenlerinin birlikte etkileşimlerine bakıldığında ise intihar olasılığı ölçeği ortalama toplam puanının da değişkenlerin temel etkileri istatistiksel olarak anlamlı bulunurken, ikili etkileşim etkileri istatistiksel olarak anlamsız bulunmuştur.

Değişkenler arasındaki ana etkiler, ikili ve daha yüksek dereceden etkileşim etkilerinin sınanması amacıyla oluşturulan logaritmik doğrusal analiz sonuçlarına göre ise doymuş hiyerarşik logaritmik doğrusal model ve doymamış logaritmik doğrusal model olmak üzere iki ayrı logaritmik doğrusal model elde edilmiştir.

“Cinsiyet”, “Sosyo Ekonomik Düzey”, ve “İntihar Olasılığı Ölçeği” değişkenlerine ilişkin elde edilen, ana etkiler ve ikili etkileşim etkilerini içeren doymamış logaritmik doğrusal model sonuçlarına göre hücre frekansına en önemli katkıyı sağlayan değerin cinsiyet değişkeninin kadın kategorisine ait olduğu ve intihar olasılığı düşük riskli grupta yer alanların yine cinsiyet değişkeninin kadın kategorisinde yer alanlara bağımlı olduğu tespit edilmiştir. Kadınların düşük riskli intihar grubunda olma oranı, erkeklere göre 2.13 kat daha fazla olarak bulunmuştur. Örnek odds oranı (2.135) kullanılarak hesaplanan %95 Güven Aralığı 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 sosyo ekonomik düzeyi yüksek olanların düşük riskli intihar grubunda olma oranının, sosyo ekonomik düzeyi düşük olanların düşük riskli intihar grubunda olma oranından 1.63 kat daha fazla bulunmuştur. Ancak örnek odds oranı kullanılarak hesaplanan %95 Güven Aralığı, değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içerdiğinden hesaplanan odds oranının istatistiksel olarak anlamlı olmadığı yani gruplar arasında farklılık olmadığı sonucuna ulaşılmıştır.

“Cinsiyet”, “Problem Çözme Envanteri” ve “Öğrenim Düzeyi” değişkenlerine ilişkin elde edilen ana etkiler, ikili etkileşim etkileri ve üçlü etkileşim etkilerini içeren doymuş hiyerarşik logaritmik doğrusal model sonuçlarına göre ise hücre frekansına en önemli katkıyı sağlayan değer cinsiyet değişkeninin kadın kategorisine aittir. Öğrenim düzeyi lise olanların problem çözme becerileri, öğrenim düzeyi üniversite olanların problem çözme becerilerinden daha yüksektir. Ancak örnek odds oranı (1.11) kullanılarak hesaplanan %95 Güven Aralığı, değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içerdiğinden hesaplanan odds oranı istatistiksel olarak anlamlı bulunmamıştır. Ayrıca üniversite öğrencilerinin düşük riskli intihar gurubunda olma oranı, lise öğrencilerinin düşük riskli intihar grubunda olma oranının 1.05 katı olarak tespit edilmiştir. Ancak örnek odds oranı (0.95) kullanılarak hesaplanan %95 Güven Aralığı, değişkenlerin birbirinden bağımsız olduğunu ifade eden $\theta = 1$ değerini içerdiğinden hesaplanan odds oranının istatiksel olarak anlamlı olmadığı dolayısıyla gruplar arasında farklılık olmadığı tespit edilmiştir.

Koşullu odds oranları sonuçlarına göre; cinsiyetin kadın olduğu biliniyorken, lise öğrencilerinin problem çözme becerileri, üniversite öğrencilerine göre daha yüksektir. Cinsiyetin erkek olduğu bilindiğinde ise; üniversite öğrencilerinin problem çözme becerileri, lise öğrencilerine göre daha yüksek olduğu tespit edilmiştir. Benzer şekilde öğrenim düzeyinin lise olduğu bilindiğinde, kadınların problem çözme becerilerinin, erkeklere göre daha yüksek olduğu tespit edilirken, öğrenim düzeyinin üniversite olduğu bilindiğinde ise, erkeklerin problem çözme becerilerinin, kadınlara göre daha yüksek olduğu tespit edilmiştir.

Özetle, bu çalışmada değişkenler arasındaki ikili ve daha yüksek dereceden etkileşimler logaritmik doğrusal modeller yöntemi ile sadece üç yönlü çapraz tablolar için incelenmiştir. Değişkenler arasındaki ilişkiler ise odds oranları ile yorumlanmıştır. Değişkenler arasındaki karmaşık ilişki yapısını ortaya çıkaran aynı anda birden fazla hipotezin test edilmesine olanak sağlayan logaritmik doğrusal modeller, çok yönlü çapraz tabloların ($k \geq 4$) analizinde de kullanılabilir. Ancak değişken sayısı arttıkça bağımsızlık eşitlikleri de buna bağlı olarak artar ve etkileşim terimlerini açıklamak karmaşık, kolay yorumlanamayan bir hal alır.

Bu tez çalışmasının 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 ilişkisel düzeyde olup, değişkenler arasında neden-sonuç ilişkisi vermemektedir.

KAYNAKLAR

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.

Agresti, A. (2002). Categorical Data Analysis, Second Edition, John Wiley & Sons, New Jersey.k

Becanım, C. (2006). *Log-Lineer Modeller ve Doktor-Tıbbi Satış Mümessilleri İlişkileri Üzerine Bir Uygulama* . Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.

Brzezinska, J. (2013). Model selection methods in log-linear analysis. *Acta Universitatis Lodziensis Folia Oeconom* , 285: 107-114.

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.

Christensen, R. (1997). *Log-linear models and logistic regression, NY, Springer, 484 pp.*

Clum GA, Patsiokas A, Luscomb R ve ark. (1979) Emperically based comprehensive treatment program for parasuicide. J Consult Clin Psychol, 47: 937-45.

Cull, JG, & Gill, WS (1988). Suicide Probability Scale (SPS) Manual. . Western Psychological Services, Los Angeles.

Çilan, A. (2013). *Sosyal Bilimlerde Kategorik Verilerle İlişki Analizi.* Pegem Akademi, 199 s.

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 adresinden "15.11.2018" tarihinde alındı.

Demirhan, H. (2004). Logaritmik Doğrusal Modellerde Parametrelerin ve Beklenen Göze Sıklıkların Bayesci Kestirimi. Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Ankara,2004.

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., & 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 qualiative data:interaction among multiple classification". J.Amer.Statist.Assos.,65,226-256 (1970).

Haberman, S. J. (1974). "Loglinear models for frequency tables with ordered classifications". Biometrics, 36:589-600 (1974).

Heppner, PP, & Petersen, CH (1982). The development and implications of a personal problem solving inventory. . Journal of Counseling Psychology, 29: 66-75.

Hisli Şahin, N., & Durak Batıgün, A. (2009). Lise ve Üniversite Öğrencilerinde İntihar Riskini Belirlemeye Yönelik Bir Modelin Sınanması. *Türk Psikiyatri Dergis* , Türk Psikiyatri Dergisi 2009; 20(1):28-36.

Howell, D. (2010). Statistical methods for psychology. Cengage Learning, 768 pp.

Jeansonne, A. (2017). *Loglinear Models*. Ocak 2, 2017 tarihinde http://userwww.sfsu.edu/efc/classes/biol710/loglinear/Log%20Linear%20Models.pdf adresinden alındı

Karabulut, E. (1998). Log Linear Modeller ve Bir Uygulaması. Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara, Türkiye.

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.

Knoke, D., & Burke, P. J. (1980). Log-Linear Models, Sage Publications, Newbury Park.

Liao, T.F; (1994) Interpreting Probability Models: Logit, Probit and Other Generalized Lineer Models, Sage Publications, Thousand Oaks.

Mete, S. (2009). Kategorik Veri Analizi Yöntemleri ve Uygulamalar,Doktora Tezi, Ekonometri Anabilim Dalı. Ankara: Gazi Üniversitesi.

Mete, S., & 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.

Oğuzlar, A. (2004). Hiyerarşik Logaritmik Doğrusal Modeller Arasından En Uygun Modelin Seçimi, Öneri Dergisi, C.6, S.21, ss.235-245.

Özdil, Ö. (2009). Kurum kültürü işlevlerinin çok yönlü kontenjans tabloları ile incelenmesi. İstanbul Üniversitesi.

Öztürk, F. (2011). *Olasılık ve İstatistiğe Giriş I.* Ankara: Gazi Kitabevi, 245 s.

Powers, D. A., & Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis.* Academic Pres.

Şahin, N., & Batıgün, AD. (2000). Yaşamı sürdürme nedenleri ve intihar olasılığı, (Yayınlanmamış Çalışma).

Şahin, N., Şahin, NH, & Heppner, PP ve ark. (1993). *Psychometric proporties of the Problem Solving Inventory in a group of Turkish university students.* Cognitive Therapy and Research, 17(4): 379-96.

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 Çözümlenmesi ve Log-Linear Modeller . *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:8, Sayı:1* , (s:299-308).

Yılmaz, V., & 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., & Erdugan, F. (2015). Kontenjans Tablolarının Analizinde Log-Lineer Modellerin Kullanımı ve Sigara Bağımlılığı Üzerine Bir Uygulama. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi Cilt 19 Sayı 2* , 221-235.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Esra TOPALOĞLU

Doğum Yeri : Urla

Doğum Tarihi : 07.12.1994

Medeni Durumu : Bekar

E-Posta : topalogluesra3@gmail.com

Eğitim Durumu

Lisans : Süleyman Demirel Üniversitesi / Ekonometri / 2012-2016