

## Kontenjans tablolarının analizinde log-linear modellerin kullanımı ve sigara bağımlılığı üzerine bir uygulama

Sevgi Yurt Öncel<sup>1\*</sup>, Funda Erdugan<sup>2</sup>

*04.02.2015 Geliş/Received, 26.04.2015 Kabul/Accepted*

### ÖZ

İki yönlü kontenjans tabloların istatistiksel analizlerinde ki-kare analizi kullanılmakta ancak daha büyük boyutlu kontenjans tabloların analizinde yetersiz kalmaktadır. Üç veya daha çok boyutlu kontenjans tablolarında ilişki yapılarının belirlenmesi için log-linear modeller kullanılabilir. Log-linear modeller yardımıyla nitel değişkenler arasındaki etkileşimler sorgulanabilmektedir. Bu çalışmada log-linear modeller incelenmiş ve sigara bağımlılığını açıklayabilmek için uygulamalar yapılmıştır.

**Anahtar Kelimeler:** FTND (nikotin bağımlılığı için fagerström testi), kategorik veri, kontenjans tabloları, ki-kare analizi, log-linear analiz, , sigara içme durumu

## Using log-linear models in the analysis of contingency tables and an application on the smoking addiction

### ABSTRACT

Two-way cross tables are used chi-square analysis of statistical analysis, but is insufficient for the analysis of larger cross tables. Log-linear model can be used to determine the structure of three or more multi-dimensional cross-table relationships. Loglinear model can be queried with the help of more interaction between qualitative variables. Log-linear models examined in this study and application is made to explain the smoking addiction

**Keywords:** categorical data, chi-square analysis, contingency tables, ftnd (fagerstrom test for nicotine dependence), log-linear models, smoking status.

---

\* Sorumlu Yazar / Corresponding Author

1 Kırıkkale Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Kırıkkale - syoncel@gmail.com

2 Kırıkkale Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Kırıkkale - ferdugan@gmail.com

## 1. GİRİŞ (INTRODUCTION)

Kategorik verinin pek çok alanda önemli bir yerinin olduğu bilinmektedir. Yatay (satur) ve düşey (sütun) eksenlerde değişkenlere ait gözlenen frekans değerleri yazılarak elde edilen tablolara kontenjans (olumsallık) tablosu denir. İlgilenilen tüm değişkenler kategorik olduğunda veriler, kontenjans tabloları ile özetlenebilir. Kategorik verilerle yapılan çalışmalarda güdülen amaç, çeşitli kontenjans tabloları düzenlemek ve düzenlenen tablolar üzerinde yapılacak analizlere dayanarak değişkenler arasındaki karmaşık ilişki yapılarını ortaya çıkartmak ve araştırılan konuyla ilgili daha doğru ve daha kapsamlı bilgi sahibi olmaktır. Bu amacı gerçekleştirmenin bir yolu da kontenjans tablosunun boyutlarını oluşturan değişkenlerin açıklanan değişken üzerindeki ana etkilerini, iki değişkenli etkileşim etkilerini ve üç değişkenli etkileşim etkilerini kestirmek, sınamak ve en uygun log-lineer modelle ifade etmektir.

Günümüzün en önemli sağlık sorunlarından birisi sigara bağımlılığıdır. Sigara tüketimi dünyada ve Türkiye’de başta solunum sistemi olmak üzere çeşitli hastalıklara neden olan önemli bir halk sağlığı sorunudur. Sigara bağımlılığı da diğer birçok bağımlılık gibi, hem genetik hem de çevresel yönleri olan kompleks bir davranıştır[1-5].

WHO Avrupa Bölge Raporuna göre, 2009 yılında, ABD’de yetişkinler arasında sigara içenlerin sayısı (erkek % 26.4, kadın % 22.3) % 24.3 olarak tahmin edilmiştir. İskandinav ve bazı Batı Avrupa ülkelerinde, kadınlar ve erkekler için sigara oranları benzer ve gittikçe azalıyor. Örneğin, erkek ve kadın sigara içenlerin oranları sırasıyla Norveç’te % 31 ve % 28, İrlanda’da % 31 ve % 26, Hollanda’da ve Finlandiya’da % 28 ve % 22’dir. Orta ve Güney Avrupa’da, erkeklerin sigara içmesi kadınlardan daha fazladır. Erkeklerin kadınlar karşı sigara içme oranları Yunanistan’da % 63’e % 41, Avusturya’da % 47’ye % 45 ve Bulgaristan’da % 48’e % 27 dir. Dünya Sağlık Örgütü (WHO) (2009)’nün Küresel Yetişkin Tütün Araştırmasına göre, tüm bireylerin (erkeklerde % 47.8, kadınlarda% 15.1) arasında Türkiye’de de düzenli sigara içme oranı 31.1 olarak tahmin edilmektedir [6-7]. Sigara bağımlılığının genetiğinde klasik Mendelyen kalıtım örüntüsü izlenmemektedir. İkizlerle yapılmış 14 farklı çalışmanın değerlendirildiği bir çalışmada nikotin bağımlılığının %60 genetik, %20 ailesel, %20 çevresel faktörlerden kaynaklandığı ifade edilmiştir. Yapılan çalışmalar sigara bağımlılığının gelişiminde; çevresel faktörlerin sigara içmeye başlama ile genetik faktörlerin ise düzenli içicilikten bağımlılığa geçişte daha belirgin bir rol oynadığını göstermiştir [3].

Bricker ve ark. [8] nın çalışmalarına göre yakın arkadaş, kardeş ve anne-babanın sigara kullanması, çocuklarının

sigara kullanması üzerindeki en büyük etkilerdi. El-Sharkawy [9] çalışmasında belirtmiştir ki cinsiyetin erkek olması sigara bağımlılığını artıran bir risk faktörüdür.

Öncel ve arkadaşları [10]’nın çalışmalarına göre sigara içen bir kişi olma riski, kadınlara göre erkeklerde 2.26 kat daha fazlaydı. Sigara içen bir kardeşe sahip olmak 1.95 kat, alkol kullanan bir kişi olmak 2.11 kat sigara içme riskini artıran faktörler olarak bulunmuştur. Aynı çalışmada, yaşları arasındaki fark 0-7 olan aynı cinsiyetten kardeşler arasında yapılan incelemelere göre sigara içen bir kardeşinin olması kadınlarda 4.7 kat, erkeklerde 5 kat sigara içme riskini artırdığı belirtilmiştir [10-11].

Sigara bağımlılığı araştırmalarında kullanılan yaygın yöntemlerden biri anket formuyla toplanan verilerden yola çıkarak kitle hakkında sonuç çıkarmaktır. Bu anket formlarında yer alan iki nitel değişken arasındaki ilişki klasik kontenjans tablo analizleri ile incelenebilir, ancak bu yöntem, ikiden fazla değişkenin ilişkisini aynı modelde inceleyebilmek için yetersiz kalmaktadır. Literatürde kategorik verilerin araştırılması amacıyla, 3 yönlü kontenjans tablolarında log-lineer modellerin kullanımı ile ilgili yapılmış birçok çalışma mevcuttur. Bu çalışmalarda genellikle ilgilenilen değişkeni etkileyen faktörler, uygun bir log-lineer model ile açıklanmaya çalışılmış ve anlamlı bulunan sonuçların yorumlanması ile ilgilenilmiştir. Bunlardan bazıları ise kısa bir şekilde özetlenmiştir. Erbaş ve İçöz [12] çalışmasında, 1994 yılı adalet istatistiklerini kullanarak modelin bir uygulamasını yapmıştır. Bülbül [13] çalışmasında, çok boyutlu çapraz tabloların çözümlenmesinde logaritmik doğrusal modellerin kullanımını, 6-17 yaş çocuk işgücüne ilişkin yerleşim yeri, çalışılan sektör ve cinsiyet kategorik değişkenlerinden oluşan üç boyutlu kontenjans tablosu üzerinde, verilere en uygun modeli belirleyerek model parametreleri ve birlikteliklerin hücre frekansları üzerindeki etkilerini incelemiştir. Altaş ve ark. [14] çalışmasında yurtdışında çalışıp Türkiye’ye dönen akademisyenler üzerinde elde edilen verilerden yararlanarak değişkenler arasındaki ilişki yapısını log-lineer analiz ile incelemiştir. Filiz [15], üniversite öğrencilerinde sigara, alkol ve nargile içme sıklığını belirlemek ve bunları etkileyen risk faktörlerini incelemek amacıyla yaptığı çalışmasında log-lineer modelleri kullanmıştır. Farklı değişkenler için, üç yönlü kontenjans tablolarında oluşturulabilecek log-lineer modeller arasında uygun modelin seçimi üzerine çalışmış ve günlük içilen sigara sayısının, kardeş sayısı ve ailenin sosyal durumundan etkilenmediği, cinsiyet, yaş, barınma şekli ve sigara cinsinden etkilendiği sonucuna ulaşmıştır. Bu çalışmada ise sigara ile yapılmış diğer çalışmalardan farklı olarak özellikle sigara bağımlılık düzeyi değişkeni ele alınmış, bu değişken ile cinsiyet ve ebeveynlerin sigara içme durumu

değişkenleri arasındaki ilişki yapısı araştırılmıştır Şıklar ve ark. [16] çalışmasında üniversite akademik personelinin iş tatminleri ve duygusal tükenmişliklerini log linear analiz ve uyum analizi ile ele almıştır. Mete ve Ünsal [17], kategorik veriler için log-linear modelleri kullanarak göç istatistikleri üzerine bir çalışma yapmışlardır. Yılmaz ve Aktaş [18], trafik kazalarını üç boyutlu kontenjans tablolarının log-linear modeller analizi ile açıklamışlardır.

Bu çalışmada öncelikle iki boyutlu daha sonra da üç boyutlu kontenjans tabloları için log-linear modellerin matematiksel yapısı tartışılmış ve son olarak da söz konusu modellerin kullanımını bir uygulama ile açıklanmıştır. Uygulamada, cinsiyet, ebeveynlerin sigara içme durumu ve nikotin bağımlılık düzeyi değişkenleri ile kurulan modellerde sigara bağımlılığı açıklanmaya çalışılmıştır.

## 2. LOG-LİNEER ANALİZ İLE İLGİLİ GENEL BİLGİLER (GENERAL INFORMATION ABOUT LOG-LINEAR ANALYSIS)

İstatistiksel veri analizinin en önemli aşamalarından birisi değişken yapılarına göre uygulanacak analiz yöntemlerinin seçilmesidir. Bu çalışmada kesikli çok değişkenli veri analizleri yapılacaktır. Nitel değişkenlerden elde edilen kategorik verilerin kontenjans tablolarda gösterimi sıkça kullanılır. Bu tabloların en basiti iki yönlü olanlardır. Bu tablolarda, analiz bir tek hipotezi test eder ve durur. Daha ileri gitmek mümkün olamaz. Genelde pratikte veriler üç veya daha fazla sayıda değişkenlerin oluşturduğu kontenjans tablolar ile ifade edilirler. Böyle çok boyutlu tabloların analizinde uygun modeller kurulurken değişkenlerin karşılıklı bağımsızlığına, koşullu bağımsızlığına ve kısmi bağımsızlığına dair hipotezler kurulur. Çok boyutlu tabloların analizinde kullanılan istatistik modeller ise log-linear modeller, lojistik modeller, probit modeller v.b dir [19-23].

İki kategorik değişken arasındaki ilişkinin incelendiği kontenjans tablo analizleri, değişken sayısı ikiden fazla olduğunda yetersiz kalmaktadır. Bu nedenle çok boyutlu tabloların analizinde kategorik değişkenlerin birbirleriyle olan etkileşimlerine ve değişkenlerin tek başına anlamlılığına dair hipotez testlerinin yapıldığı log-linear analizler kesikli çok değişkenli analizler arasında önemli bir yere sahiptir. Kontenjans tablolarında log-linear analizi, sürekli dağılıma sahip faktörün yanıt değişkenleri ile açıkladığı, varyans analizine (ANOVA) benzerdir. Log-linear analiz, en iyi modelin bulunmasında ve kurulan modelin anlamlılığının test edilmesinde de kullanılır [22]. Bu çalışmada en iyi modelin elde edilmesi konusu işlenmemiştir. Öncel ve arkadaşları [10-

11]'nın çalışmalarında anlamlı bulunan değişkenler için log-linear analiz yapılarak model parametreleri ve etkileşim terimleri için parametre tahmini ve anlamlılıkları araştırılmıştır.

Genel log-linear model, kontenjans tablo içindeki her bir hücre frekansının logaritmasını, tabloda mevcut olan değişkenler arasındaki mümkün her etkileşimin bir doğrusal kombinasyonu olarak ifade eder. Kategorik değişkenlerden oluşan verilerin araştırılmasında ve bir değişkenin ya da değişkenlerin kombinasyonunun etkisinin gösterilmesinde, veri yapılarının bileşen parçalarına ayrılmasında yararlıdır. Ayrıca log-linear analizi, parametre tahminlerinin elde edilmesini sağlamanın yanı sıra araştırmacıya karmaşık bir çoklu kontenjans tablonun içerdiği bilgiyi kısaltma olanağı da verir. Yani verinin daha sade bir formla temsil edilip edilemeyeceğini gösterir [21]. Log-linear modelin matematiksel yönünü tartışmak için öncelikle boyutlu iki yönlü, sonrasında ise  $I \times J \times K$  boyutlu üç yönlü kontenjans tabloları ele alınmıştır.

### 2.1. İki Yönlü Kontenjans Tablolarında Log-Linear Analiz (Log-Linear Analysis of Two-Way Contingency Tables)

İki yönlü ve  $I \times J$  boyutlu tabloların genel hali için log-linear modeli göz önüne alınsın. Satırda yer alan  $i = 1, \dots, I$  düzeyli kategorik değişken  $A$  ile, sütunda yer alan  $j = 1, \dots, J$  düzeyli kategorik değişken ise  $B$  ile gösterilsin. Bu durumda kontenjans tablosu

Tablo 1. İki yönlü  $I \times J$  boyutlu log-linear model (Two-way  $I \times J$  dimensional log-linear model)

		2. Değişken ( $B$ )				Toplam
		1	2	...	$J$	
Değişken ( $A$ )	1	$f_{11}$	$f_{12}$	...	$f_{1J}$	$f_{1+}$
	2	$f_{21}$	$f_{22}$	...	$f_{2J}$	$f_{2+}$
	...	...	...	...	...	...
	$I$	$f_{I1}$	$f_{I2}$	...	$f_{IJ}$	$f_{I+}$
Toplam		$f_{+1}$	$f_{+2}$	...	$f_{+J}$	$f_{++}$

şeklinde olacaktır. Burada  $f_{ij}$ ,  $A$  değişkeninin  $i$ .inci düzeyi ile  $B$  değişkeninin  $j$ .inci düzeyinde yer alan hücrenin gözlenen frekans değeridir.  $l_{ij} = \ln(f_{ij})$  olmak üzere,  $I \times J$  boyutlu tabloların doymuş (saturated) log-linear modeli

$$l_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, i=1, \dots, I; j=1, \dots, J \quad (1)$$

biçimindedir. Burada sırasıyla,  $\mu$  frekans tablosundaki tüm gözelerin logaritmalarının ortalama değeri,  $\lambda_i^A$ , A değişkeninin  $i$ . inci düzeyinin ana etki parametresi;  $\lambda_j^B$ , B değişkeninin  $j$ . inci düzeyinin ana etki parametresi ve  $\lambda_{ij}^{AB}$ , AB değişkenlerinin etkileşimlerinin  $i$ . ve  $j$ . kombinasyon düzeyinin etki parametresini göstermektedir. Sabit etkili ANOVA modeline paralel olarak log-lineer modelde de değişkenlerin kategorilerine ait katsayıların toplamları sıfıra eşittir. Yani model için kısıtlar

$$\sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0 \quad (2)$$

şeklindedir. Model, tablodaki hücrelerin toplam sayısı yani  $I \times J$  kadar parametreye sahiptir. Bu parametrelere ait tahminler ise genel frekans toplamı  $l_{++} = \sum_{ij} \ln(f_{ij})$  marjinal frekans toplamları  $l_{i+} = \sum_j \ln(f_{ij})$  ve  $l_{+j} = \sum_i \ln(f_{ij})$  olmak üzere log-lineer modelin parametrelerinin tahmin edicileri Tablo 2’de verildiği gibidir. Ayrıca, bu tabloda log-lineer model ile ANOVA arasındaki benzerliğe dikkat çekmek için ANOVA etkilerine ait tahmin ediciler de verilmiştir.

Gruplar arası farklılığın incelendiği ANOVA modellerinde deneme etkileri ve etkileşimlerinin yer aldığı model

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk} \quad (3)$$

biçimindedir. Varyans analizinde deneme etkilerini hesaplarırken deneme ve blok ortalamaları etrafındaki değişim göz önünde bulundurulur. Varyans analizi

modellerinde denemeler kategorik değişkenlerdir ve bu değişkenler sınıflandırılmış veya sıralanmış olabilir. Bu deneme düzeylerinde ölçme ile elde edilmiş verinin normal dağılıma sahip olması ANOVA’nın temel varsayımdır. (3) eşitliği ile gösterilen model iki yönlü, etkileşimli varyans analizi modelidir. Eğer elimizde kategorik değişkenlerin her bir düzeyine ait frekans verisi varsa o zaman normal dağılım varsayımı gerektirmeyen log-lineer model kullanımı uygundur. Log-lineer modelde frekansların kendisi yerine doğal logaritmaları alınarak analiz yapılır. Varyans analizinde  $\mu$ ,  $\alpha_i$ ,  $\beta_j$  ve  $\alpha\beta_{ij}$  terimleri ile gösterilen deneme etkileri, log-lineer modellerde  $\lambda$ ,  $\lambda_i^A$ ,  $\lambda_j^B$  ve  $\lambda_{ij}^{AB}$  ile gösterilir.

Tablo 2. İki yönlü log-lineer model ve ANOVA modelinde tahmin ediciler (Estimators in two-way log-linear model and ANOVA model)

<b>Log-lineer model etkileri</b>
$\hat{\lambda} = \frac{l_{++}}{IJ}$
$\hat{\lambda}_i^A = \frac{l_{i+}}{J} - \frac{l_{++}}{IJ}$
$\hat{\lambda}_j^B = \frac{l_{+j}}{I} - \frac{l_{++}}{IJ}$
$\hat{\lambda}_{ij}^{AB} = l_{ij} - \frac{l_{i+}}{J} - \frac{l_{+j}}{I} + \frac{l_{++}}{IJ} = l_{ij} - (\hat{\mu} + \hat{\lambda}_i^A + \hat{\lambda}_j^B)$
<b>ANOVA etkileri</b>
$\hat{\mu} = \bar{X}_{..}$
$\hat{\alpha}_i = \bar{X}_{i.} - \bar{X}_{..}$
$\hat{\beta}_j = \bar{X}_{.j} - \bar{X}_{..}$
$\hat{\alpha}\hat{\beta}_{ij} = \bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..}$

Parametrelerin anlamlılığını araştırmak için Tablo 2’de verilen tahmin edicilerin asimptotik varyansları,

$$\begin{aligned}
 Var(\hat{\lambda}_r^A) &= \frac{1}{(IJK)^2} \left[ \gamma^2 \sum_{j=1}^J \sum_{k=1}^K \frac{1}{f_{rjk}} + \sum_{i \neq r}^I \sum_{j=1}^J \sum_{k=1}^K \frac{1}{f_{ijk}} \right] \\
 Var(\hat{\lambda}_c^B) &= \frac{1}{(IJK)^2} \left[ \beta^2 \sum_{i=1}^I \sum_{k=1}^K \frac{1}{f_{ick}} + \sum_{i=1}^I \sum_{j \neq c}^J \sum_{k=1}^K \frac{1}{f_{ijk}} \right] \\
 Var(\hat{\lambda}_s^C) &= \frac{1}{(IJK)^2} \left[ \delta^2 \sum_{i=1}^I \sum_{j=1}^J \frac{1}{f_{ijs}} + \sum_{i=1}^I \sum_{j=1}^J \sum_{k \neq s}^K \frac{1}{f_{ijk}} \right] \\
 Var(\hat{\lambda}_{rc}^{AB}) &= \frac{1}{(IJ)^2} \left[ (\gamma\beta)^2 \frac{1}{f_{rc}} + \gamma^2 \sum_{j \neq c}^J \frac{1}{f_{rj}} + \beta^2 \sum_{i \neq r}^I \frac{1}{f_{ic}} + \sum_{i \neq r}^I \sum_{j \neq r}^J \frac{1}{f_{ij}} \right] \\
 Var(\hat{\lambda}_{rcs}^{ABC}) &= \frac{1}{(IJK)^2} \left[ (\gamma\beta\delta)^2 \frac{1}{f_{rcs}} + (\beta\delta)^2 \sum_{i \neq r}^I \frac{1}{f_{ics}} + (\gamma\delta)^2 \sum_{j \neq c}^J \frac{1}{f_{rjs}} + (\gamma\beta)^2 \sum_{k \neq s}^K \frac{1}{f_{rck}} \right. \\
 &\quad \left. + \gamma^2 \sum_{j \neq c}^J \sum_{k \neq s}^K \frac{1}{f_{rjk}} + \beta^2 \sum_{i \neq r}^I \sum_{k \neq s}^K \frac{1}{f_{ick}} + \delta^2 \sum_{i \neq r}^I \sum_{j \neq c}^J \frac{1}{f_{ijs}} + \sum_{i \neq r}^I \sum_{j \neq c}^J \sum_{k \neq s}^K \frac{1}{f_{ijk}} \right]
 \end{aligned} \tag{4}$$

formülleri ile hesaplanır. Burada  $\gamma = I - 1$ ,  $\beta = J - 1$ ,  $\delta = K - 1$ ,  $r = 1, \dots, I$ ,  $c = 1, \dots, J$ ,  $s = 1, \dots, K$  dır.

(4) eşitliğinde yer alan formüller, diğer ikili ve üçlü etkileşimler için benzer şekildedir. Bu parametrelerin anlamlılığını test etmek için

$$Z(\hat{\lambda}) = \frac{\hat{\lambda}}{\sqrt{Var(\hat{\lambda})}} \tag{5}$$

(5) formulu ile verilen test istatistiğinin hesaplanan değeri,  $Z_\alpha$  (standart normal dağılım tablosundan  $\alpha$  'lık alana karşılık gelen değer) tablo değeri ile karşılaştırarak karar verilir. Hesaplanan değer mutlak değeri, tablo değerinden büyükse ilgilenilen parametrenin anlamlı olduğu sonucuna varılarak model-denkleminde bu parametreye yer verilir [23]. Şimdiye kadar verilerin modele iyi uyduğunu varsayarak log-lineer modellerle elde edilen parametre tahminlerini inceledik. Gerçekte test edilmesi gereken bir başka nokta da log-lineer modelin gözlem değerlerine iyi uyup uymadığıdır. Bu işlem önerilen model için beklenen değerlerin tahmin edilmesini ve bu değerlerin gözlenen değerlerle Pearson  $\chi^2$  veya olabilirlik oranı (likelihood ratio)  $G^2$  testleri kullanılarak karşılaştırılmasını içerir. Yani belirli bir modelin verilere uyup uymadığını gösteren hipotezin testi,  $f_{ij}$ : gözlenen değer,  $F_{ij}$ : beklenen değer olmak üzere

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - F_{ij})^2}{F_{ij}} \tag{6}$$

formülü ile Pearson  $\chi^2$  istatistiği temeline dayanır.

Genelde  $G^2$  istatistiği  $\chi^2$ 'ye göre daha çok tercih edilir.

$G^2$  testinin genel gösterimi

$$G^2 = 2f_{ij} \ln \left( \frac{f_{ij}}{F_{ij}} \right) \tag{7}$$

biçimindedir. Kullandığımız model doğru ise hem  $\chi^2$

hem de  $G^2$  testleri  $(I-1)(J-1)$  serbestlik dereceli  $\chi^2$  dağılımına sahiptir [22]. Uygulama olarak 2 yönlü  $3 \times 3$  tipindeki bir kontenjans tablosu için kurulan log-lineer modelin parametre tahminleri ve yorumları Bölüm 3.1 de ayrıntılı bir şekilde verilmiştir.

## 2.2 Üç Yönlü Kontenjans Tablolarında Log-Linear Analiz (Log-Linear Analysis of Three-Way Contingency Tables)

Sırasıyla  $i = 1, 2, \dots, I$ ;  $j = 1, 2, \dots, J$ ;  $k = 1, 2, \dots, K$

indisli  $A, B, C$  değişkenleri ile gösterilen üç yönlü

$I \times J \times K$  boyutlu bir kontenjans tablosu için doymuş

bir log-lineer model

$$l_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \tag{8}$$

ve kontenjans tablosu Tablo 3 ile verilmiştir. (2) eşitliği ile verilen kısıtlar bu modelde üç değişken için de geçerlidir. Tabloda  $i = 1, 2, \dots, I$ ;  $j = 1, 2, \dots, J$ ;  $k = 1, 2, \dots, K$  ilgili değişkenlerin kategorik düzeyini göstermek üzere,  $(i, j, k)$  gözesinin gözlenen frekansı  $f_{ijk}$  ile gösterilmiştir. Bu model için ana etki, ikili ve üçlü



Bu çalışma için 745 bireye anket yapıldı. Bunlardan 447 (%60,1)'si erkek, 297 (%39,9)'si kadındı. Anket sonuçlarına göre, erkeklerden 290 kişi (%64,9), kadınlardan 127 kişi (%42,8) bir ay veya daha uzun süreden beri her gün sigara içen kişilerdi. Ailesinden kimse sigara içmemesine rağmen 103 (%42,4) kişi sigara içen kişilerdi. Sigara kullanan bireylerin yaşı, ne kadar süredir sigara içtiği (ay), günlük içilen sigara sayısı ve ftn skoru için hesaplanan ortalama (standart sapma), erkeklerde sırasıyla 35 (11), 148 (125), 18 (9), 4.56 (2,29) kadınlarda ise 30 (10), 98 (89), 13 (7), 3.53 (2,72) olarak bulunmuştur. Sigara içen bireylerin cinsiyete göre sigara içmeye başlama nedeni olarak erkeklerde arkadaş etkisi %44,1, merak %28,6, özentisi %27,2, stres %22,4, ortama uyum sağlama %14,8, ailenin diğer bireylerinin de sigara tüketmesi %7,9, filmler %3,4 ve reklamların etkisi %1,4 iken kadınlarda arkadaş etkisi %48, merak %34,6, özentisi %22,8, stres %38,6, ortama uyum sağlama %15, ailenin diğer bireylerinin de sigara tüketmesi %7,1, filmler %0,8 ve reklamların etkisi %0,8 olarak bulunmuştur. Erkeklerde ilk sigaraya başlama yaşı en az 11 iken kadınlarda bu değer 15'dir. Ortalama erkekler 18 yaşında sigaraya başlarken kadınların 20 yaşında başladığı görülmüştür. Anne veya babanın ikisinin de sigara içmemesi veya her ikisinden yalnız birinin sigara içmesi veya her ikisinin de sigara içmesi şeklinde kategorik değişken olarak alınan ebeveynlerin sigara içmesi durumu ile bireyin sigara içmesi veya içmemesi durumunun birlikte değişiminin incelenmesi için kontenjans tablo analizi yapılmıştır. Analiz sonuçlarına göre Ki-kare hesap değeri 15,261 ( $p < 0,05$ ) olarak bulunmuştur. Buna göre bireyin sigara içmesi ebeveynlerinin sigara içmesinden bağımsız değildir, aralarındaki ilişki ise 0,265 (Gamma değeri) dir.

### 3.1 İki Yönlü Kontenjans Tablolarında Analiz (Analysis of two way contingency tables)

#### 3.1.1 Klasik Ki-kare Analizleri (Classical Chi-square Analysis)

Çalışmada kullanılacak olan değişken isimleri ve değişkenlere ait düzeyler ise Tablo 5'de verilmiştir. Burada FTND2, az bağımlı (0-3), orta bağımlı (4-6) ve çok bağımlı (7+) olacak şekilde üç düzeyli değişken; ebeveynin sigara içme durumu, ebeveynlerin hiç biri sigara içmiyor, ebeveynlerden yalnızca birisi sigara içiyor ve her ikisi de sigara içiyor şeklinde üç düzeyli değişken; cinsiyet ise kadın, erkek olacak şekilde iki düzeyli değişkendir. Çalışmanın bundan sonraki bölümlerinde (1)-(8) eşitlikleriyle verilen formüllerde kullanım kolaylığı olması bakımından, bu değişkenler sırasıyla A, B, C isimleriyle anılacaktır.

Tablo 5. Değişkenlere ait düzeyler (Levels of variables)

Risk Faktörleri	Değişken isimleri	1.Düzye	2.Düzye	3.Düzye
Cinsiyet	C	Kadın	Erkek	-
Ebeveynin sigara içme durumu	B	ikisi de içmiyor	biri içiyor	ikisi de içiyor
FTND2	A	0-3	4-6	7+

Bu değişkenlere ilişkin kontenjans tablosu Tablo 6'da verilmiştir. Bu kesimde, iki yönlü kontenjans tablosunun değerlendirilebilmesi için akla gelen ilk klasik analizler; bağımsızlık testi, dağılıma uyum iyiliği analizi, koşullu bağımsızlık testleri uygulamalı ele alınacaktır. Sadece A ve B değişkenlerinin oluşturduğu kontenjans tablosu olmak üzere " $H_{01}$ : A ile B değişkenleri birbirinden bağımsızdır", " $H_{11}$ : A ile B değişkenleri birbirinden bağımsız değildir" hipotezlerini test etmek için gözlenen ( $f$ ) ve beklenen (teorik) ( $F$ ) frekanslar, Tablo 7' de verilmiştir.

Tablo 6. A ve B değişkenleri için iki yönlü kontenjans tablosu (Two-way contingency tables for A and B variables)

		B		
		İkisi de içmiyor	Biri içiyor	İkisi de içiyor
A	0-3	53	30	12
	4-6	69	52	20
	7+	29	24	12

Tablo 7. Bağımsızlık durumunda değişkenlerin gözlenen ve beklenen frekansları (Observed and expected frequencies of the variables in case of independence)

		B			Toplam	
		1	2	3		
A	1	$f_{1j}$	53	30	12	95
		$F_{1j}$	47,7	33,5	13,9	95,0
	2	$f_{2j}$	69	52	20	141
		$F_{2j}$	70,7	49,7	20,6	141,0
	3	$f_{3j}$	29	24	12	65
		$F_{3j}$	32,6	22,9	9,5	65,0
Toplam	$f_{+j}$	151	106	44	301	

$$(7) \text{ eşitliği ile verilen } G^2 = 2f_{ij} \ln \left( \frac{f_{ij}}{F_{ij}} \right) = 2,255263$$

olarak hesaplanmıştır. Bu değer  $(3-1)(3-1)=4$  serbestlik dereceli (sd)'li ki-kare tablo değeri ile karşılaştırıldığında  $\chi^2_{(0,05;4)} = 9,49 > 2,255263$  olduğundan  $H_{01}$  hipotezi red edilemez.

Yani FTND2 değişkeni ile ebeveynlerin sigara içme durumu değişkenleri birbirinden bağımsız değişkenlerdir. Bu iki değişken arasında bir ilişki yoktur. Literatürde yer alan pek çok çalışmada yer alan genetik araştırmalarla elde ettiğimiz bu sonucun ters düşmesi nedeniyle daha ileri analizlerin yapılmasına gerek görülmüştür [3,6,7,10].

Bu şekilde elde edilmiş frekans verisi için dağılıma uyum iyiliği analizi için kurulan hipotez ise " $H_{02}$  : Sigara içen bireylerin 9 göze dağılımı rasgeledir.", " $H_{12}$  : Sigara içen bireylerin 9 göze dağılımı rasgele değildir." biçimindedir.  $H_{02}$  hipotezinin doğru olduğu varsayımı altında, beklenen frekanslar  $\frac{301}{9} = 33,44$  biçiminde olacaktır.

Tablo 8. Dağılıma uyum iyiliği analizinde değişkenlerin gözlenen ve beklenen frekansları (The observed and expected frequencies of the variables in goodness of fit analysis)

		B			Toplam	
		1	2	3		
A	1	$f_{1j}$	53	30	12	95
		$F_{1j}$	33,4	33,4	33,4	
	2	$f_{1j}$	69	52	20	141
		$F_{1j}$	33,4	33,4	33,4	
	3	$f_{1j}$	29	24	12	65
		$F_{1j}$	33,4	33,4	33,4	
Toplam		$f_{+j}$	151	106	44	301
		$F_{+j}$	151	106	44	

verilmiştir. (7) eşitliği ile verilen  $G^2 = 94,96273$  dir. Bu test istatistiği değeri  $(9-1)=8$  serbestlik dereceli ki-kare tablo değeri ile karşılaştırılır.  $\chi^2_{(0,05;8)} = 15,5 < 94,96273$  olduğundan  $H_{02}$  hipotezi red edilir. Yani, sigara içen bireylerin 9 göze dağılımı rasgele değildir.

Tablo 6'daki gibi elde edilmiş frekans verisi için FTND2 düzeyleri arasında fark olmadığını araştıran koşullu

bağımsızlık analizi için kurulan hipotez ise " $H_{03}$  : Sigara içen bireylerin FTND2 düzeyleri arasında anlamlı bir fark yoktur.", " $H_{13}$  : Sigara içen bireylerin FTND2 düzeyleri arasında anlamlı bir fark vardır." biçimindedir ve  $H_{03}$  hipotezinin doğru olduğu varsayımı altında koşullu bağımsızlık için beklenen frekanslar Tablo 9 da verilmiştir.

Tablo 9. Birinci koşullu bağımsızlık durumunda değişkenlerin beklenen frekansları (The expected frequencies of the variables in the first case of conditional independence)

		B			Toplam	
		1	2	3		
A	1	$f_{1j}$	53	30	12	95
		$F_{1j}$	50,3	35,3	14,7	95
	2	$f_{1j}$	69	52	20	141
		$F_{1j}$	50,3	35,3	14,7	141
	3	$f_{1j}$	29	24	12	65
		$F_{1j}$	50,33	35,33	14,7	65
Toplam		151	106	44	301	

(7) eşitliği ile verilen  $G^2 = 31,72672$  dir. Bu test istatistiği değeri 6 serbestlik dereceli ki-kare tablo değeri ile karşılaştırılır.  $\chi^2_{(0,05;6)} = 12,6 < 31,72672$  olduğundan  $H_{03}$  hipotezi red edilir. Yani çalışmamızda yer alan sigara içen bireylerin FTND2 düzeyleri arasında anlamlı bir fark vardır.

Tablo 6'daki gibi elde edilmiş frekans verisi için ebeveynlerin sigara içme durumları arasında anlamlı bir farklılık olup olmadığını araştıran koşullu bağımsızlık analizi için kurulan hipotez ise " $H_{04}$  : Sigara içen bireylerin ebeveynin sigara içme durumu arasında anlamlı bir fark yoktur.", " $H_{14}$  : Sigara içen bireylerin ebeveynin sigara içme durumu arasında anlamlı bir fark vardır" biçimindedir.

Ayrıca  $H_{04}$  hipotezinin doğru olduğu varsayımı altında koşullu bağımsızlık için beklenen frekanslar Tablo 10'da verilmiştir.



Tablo 10. İkinci koşullu bağımsızlık durumunda değişkenlerin gözlenen ve beklenen frekansları (The expected frequencies of the variables in the second case of conditional independence)

		B			Toplam	
		1	2	3		
A	1	$f_{1j}$	53	30	12	95
		$F_{1j}$	31,7	31,7	31,7	
	2	$f_{1j}$	69	52	20	141
		$F_{1j}$	47	47	47	
	3	$f_{1j}$	29	24	12	65

$F_{1j}$	21,7	21,7	21,7	
Toplam	151	106	44	301

(7) eşitliği ile verilen  $G^2 = 64,621$  'dir. Bu test istatistiği değeri 6 serbestlik dereceli ki-kare tablo değeri ile karşılaştırılır.  $\chi^2_{(0,05;6)} = 12,6 < 64,621$  olduğundan  $H_{04}$  hipotezi red edilir. Yani ebeveynlerin sigara içmeleri arasında anlamlı bir ilişki vardır. Çevresel bir etkileşim olarak birbirlerinden etkilenerek sigara içen kişilerdir. Elde edilen sonuçlar aşağıda Tablo 11'de özetlenmiştir.

Tablo 11. Modellerin karşılaştırılması (Comparison of the models)

Model No	Model Adı	Model	$G^2$	s.d.	$\chi^2_{0,05} (sd)$	P-değeri
1	Bağımsızlık	$\ln(F_{ij}) = \lambda + \lambda_i^A + \lambda_j^B$	2,255263	4	9,49	> 0,05
2	Uyum iyiliği	$\ln(F_{ij}) = \lambda$	94,96273	8	15,5	< 0,01
3	1. koşullu bağımsızlık	$\ln(F_{ij}) = \lambda + \lambda_j^B$	31,72672	6	12,6	< 0,01
4	2. koşullu bağımsızlık	$\ln(F_{ij}) = \lambda + \lambda_i^A$	64,621	6	12,6	< 0,01
5	Doymuş	$\ln(F_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	0	0		

$G^2$  değerinin, uygun serbestlik dereceli (sd)  $\chi^2_{tablo}$  değerinden büyük olması  $H_0$  hipotezinin red edilmesine yol açar ( $p < \alpha = 0,05$ ). Yani modelin verilere uymadığına karar verilir. Bu durumda yorumumuz iki değişkenin birbirinden bağımsız olmadığı aralarında anlamlı bir ilişkinin olduğu şeklindedir. Tablodaki modeller incelendiğinde sigara içme düzeyi ile ebeveynlerin sigara kullanma durumları arasındaki ilişkinin anlamlı olduğunu gösterirken Model 1, anlamlı çıkmamıştır. Dahası, kurduğumuz Model 3 ve Model 4'ün analizinde kurduğumuz hipotez testinden gelen 1. tip hata (testin anlam düzeyi)  $2\alpha$  kadar oldu. Bu problemden kaçmak için sabit terim, FTND etkisi, ebeveynin sigara içme durumunun etkisi ve bu değişkenlerin etkileşimlerinin yer aldığı model Tablo 11'de Model 5 olarak ifade edilmiştir. Aynı zamanda Model 5 ile bir kişinin sigara içmesini çevresel faktörler tetiklerken bağımlılığın düzeyi genetik faktörlerle açıklanmaktadır iddiası da araştırılabilecektir.

Bu nedenle bilinen basit kontenjans tablo analizlerinin yetersizliğinden yola çıkarak Model 5 ile ifade edilen log-lineer modellerle çalışmanın önemi görülmektedir. Aşağıdaki kesimde bu tür modeller için uygulamalar yer almaktadır.

### 3.1.2. İki yönlü kontenjans tablosu için log-lineer analiz uygulamaları (Log-linear analysis applications for a two-way contingency table)

Sabit terim, sigara bağımlılık düzeyi, ebeveynin sigara içme durumları ve her iki değişkenin etkileşimlerinin yer aldığı doymuş log-lineer model S biçimindedir. Doymuş modelin olabilirlik değeri ve serbestlik derecesi başlangıçta sıfırdır.

Aşağıda iki kategorik değişkenin etkileşim terimlerinin yer aldığı log-lineer model ile kontenjans tablolarının detaylı analizleri, model parametrelerinin tahmini ve anlamlılığı detaylı olarak elle çözümleri ile araştırılmıştır. Bu sayede okuyucuya log-lineerin analizinde parametre tahmini hakkında detaylı bilgiler verilmiş olup, SPSS paket programında analiz yapmanın zayıf yönleri ortaya konmuştur.

Doymuş modelin parametre tahmini için Tablo 6'da verilen frekansların Tablo 12'de doğal logaritma değerleri verilmiştir. Bu logaritma değerleri  $l_{ij}$  olmak üzere bunlara ait satır, sütun toplamları, ortalamaları ve genel ortalama da Tablo 12'de verilmiştir.

Tablo 12. Frekansların logaritma değerleri (Logarithm values of frequencies)

		B			$l_{i+}$	$\bar{l}_{i+}$
		ikiside içmiyor	biri içiyor	ikiside içiyor		
A	0-3	3,9703	3,4012	2,485	9,856	3,285
	4-6	4,2341	3,9512	2,995	11,181	3,727
	7+	3,3673	3,1781	2,4849	9,0303	3,010
$l_{+j}$		11,5717	10,5305	7,9655	30,06	-
$\bar{l}_{+j}$		3,8572	3,5102	2,655	-	3,340

Eşitlik (1) ile verilen modelin parametre tahmin değerleri, Bölüm 2.1'de yer alan Tablo 2'de verilen formüller yardımıyla Tablo 12'deki değerler

$$l_{++} = 3,9703 + 3,4012 + 2,4849 \\ + 4,2341 + 3,9512 + 2,9957 \\ + 3,3673 + 3,1781 + 2,4849 \\ = 30,0677$$

$$l_{1+} = 3,9703 + 3,4012 + 2,4849 = 9,8564$$

$$l_{2+} = 4,2341 + 3,9512 + 2,9957 = 11,181$$

$$l_{3+} = 3,3673 + 3,1781 + 2,4849 = 9,0303$$

ve

$$l_{+1} = 3,9703 + 4,2341 + 3,3673 = 11,5717$$

$$l_{+2} = 3,4012 + 3,9512 + 3,1781 = 10,5305$$

$$l_{+3} = 2,4849 + 2,9957 + 2,4849 = 7,9655$$

olarak hesaplanır ve aşağıda doymuş modelin parametrelerinin tahminleri

$$\hat{\lambda} = \frac{30,0677}{9} = 3,3409$$

$$\hat{\lambda}_1^A = \frac{l_{1+}}{J} - \frac{l_{++}}{IJ} = \frac{9,8564}{3} - \frac{30,0677}{9} = -0,0554$$

$$\hat{\lambda}_1^B = \frac{l_{+1}}{J} - \frac{l_{++}}{IJ} = \frac{11,5717}{3} - \frac{30,0677}{9} = 0,5163$$

$$\hat{\lambda}_{11}^{AB} = l_{ij} - \frac{l_{i+}}{J} - \frac{l_{+j}}{I} + \frac{l_{++}}{IJ} \\ = 3,9703 - 3,2855 - 3,8572 + 3,3409 \\ = 0,1685$$

veya

$$\hat{\lambda}_{11}^{AB} = l_{11} - \hat{\mu} - \hat{\lambda}_1^A - \hat{\lambda}_1^B \\ = 3,9703 - 3,3409 + 0,0554 - 0,5163 \\ = 0,1685$$

formülüyle elde edilir. Elde edilen bu tahmin değerleri her iki değişkeninin de ilk düzeyleri için modelde yerine yazılırsa

$$l_{11} = \mu + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB} \\ = 3,3409 - 0,0554 + 0,5163 + 0,1685 \\ = 3,9703$$

buradan ise  $e^{3,9703} = 53$  olduğu görülebilir. Tablo 2 ile verilen parametre tahmin edicilerinden yararlanarak (1) modelinde yer alan parametreler için bulunan tahmin değerleri Tablo 13'de verilmiştir. Bu tablo ana etkiler ile ikili etkileşimlere ait tahmin değerlerini içermektedir. Bağımlılığın az olduğu düzeye ait ana etki -0,0554 iken ebeveynin her ikisinin de sigara içmediği düzeye ait ana etki 0,5164 dir. Ayrıca az bağımlı ve ebeveynin sigara içmediği düzey için etkileşim değeri 0,1685 olup çok bağımlı ve ebeveynin ikisinin de sigara içtiği düzey için etkileşim değeri 0,1605 dir.

Tablo 13. Parametre tahmin değerleri (Estimations of the parameter)

		B			
		ikiside içmiyor	biri içiyor	ikiside içiyor	
A	0-3	0,1685	-0,0536	-0,1149	-0,055
	4-6	-0,0093	0,3870	-0,0456	0,3862
	7+	-0,1592	-0,0013	0,1605	-0,331
		0,5164	0,1693	-0,6857	3,3409

Kurulan modeldeki parametrelerin anlamlılığını  $\alpha = 0,05$  anlamlılık düzeyinde test etmek için  $H_0 : \lambda = 0$   $H_1 : \lambda \neq 0$  hipotezlerini göz önünde bulundurulur. Yukarıda (4) eşitliği ile verilen formül yardımıyla  $Var(\hat{\lambda}^A) = 0,047343$  olarak hesaplanmıştır.

Tablo 14. Parametrelerin önemlilik testi sonuçları (Significance test results of parameters)

Hipotez	Test istatistiği	Karar
$H_0 : \lambda^A = 0$ $H_1 : \lambda^A \neq 0$	$Z(A) = -0,25461$	Ho red edilemez
$H_0 : \lambda^B = 0$ $H_1 : \lambda^B \neq 0$	$Z(B) = 2,372881$	Ho red edilir
$H_0 : \lambda^{AB} = 0$ $H_1 : \lambda^{AB} \neq 0$	$Z(AB) = 0,774415$	Ho red edilemez

Tablo 14'den görülebileceği gibi FTND ve ebeveynlerin sigara içme durumları arasında etkileşim olmadığı ortaya çıktı. Bu durum [2-3] çalışmalarında verilen sonuçla çeliştiği için modele üçüncü bir değişken olan cinsiyet değişkeni de eklenerek 3 yönlü kontenjans tablosuna log-linear analiz uygulanacaktır.

### 3.2 Üç Yönlü Kontenjans Tablolarında Log-linear Analiz Uygulaması (Log-linear Analysis Applications in Three-Way Contingency Tables)

Üç yönlü kontenjans tablosu Tablo 15' de verilmiştir. Her bir hücre içerisinde gözlenen frekans ile bu frekansların logaritma değerleri

Tablo 15. A, B, C değişkenleri için üç yönlü kontenjans tablosu (The three-way contingency tables for A, B, C variables)

		C		
		1	2	
		$f_{ijk} (\ln(f_{ijk}))$	$f_{ijk} (\ln(f_{ijk}))$	
A	B	0	33 (3,4965)	20 (2,9957)
		1	23 (3,1355)	7 (1,9459)
		2	4 (1,3863)	8 (2,0794)
	2	0	57 (4,0431)	12 (2,4849)
		1	42 (3,7377)	10 (2,3026)
		2	16 (2,7726)	4 (1,3863)
3	0	21 (3,0445)	8 (2,0794)	
	1	21 (3,0445)	3 (1,0986)	
	2	6 (1,7918)	6 (1,7918)	

olarak hesaplanmıştır. Bölüm 2'de açıklanan ve Tablo 4'de yer alan formüllerle A, B, C terimlerine ait parametre tahmin değerleri Tablo 16.1'de,

Tablo 16.1 Ana etki parametrelerinin tahminleri (Estimations of main effects parameters)

A	B	C
$\lambda_1^A = 0,0278$	$\lambda_1^B = 0,5453$	$\lambda_1^C = 0,4604$
$\lambda_2^A = 0,3091$	$\lambda_2^B = 0,0654$	$\lambda_2^C = -0,4604$
$\lambda_3^A = -0,3370$	$\lambda_3^B = -0,6107$	

AB ikili etkileşim terimine ait parametre tahmin değerleri Tablo 16.2'de

Tablo 16.2 İkili etkileşim parametrelerinin tahminleri (Estimations of the binary interaction parameters)

		B		
		1	2	3
1	A	$\lambda_{11}^{AB} = 0,1943$	$\lambda_{12}^{AB} = -0,0313$	$\lambda_{13}^{AB} = -0,1630$
		$\lambda_{21}^{AB} = -0,0692$	$\lambda_{22}^{AB} = 0,1669$	$\lambda_{23}^{AB} = -0,0977$
2	A	$\lambda_{31}^{AB} = -0,1251$	$\lambda_{32}^{AB} = -0,1356$	$\lambda_{33}^{AB} = 0,2607$

AC ikili etkileşim terimine ait parametre tahmin değerleri Tablo 16.3'de

Tablo 16.3 İkili etkileşim parametrelerinin tahminleri (Estimations of the binary interaction parameters)

		C	
		1	2
1	A	$\lambda_{11}^{AC} = -0,2942$	$\lambda_{12}^{AC} = 0,2942$
		$\lambda_{21}^{AC} = 0,2695$	$\lambda_{22}^{AC} = -0,2695$
		$\lambda_{31}^{AC} = 0,0247$	$\lambda_{32}^{AC} = -0,0247$

BC ikili etkileşim terimine ait parametre tahmin değerleri Tablo 16.4'de

Tablo 16.4 İkili etkileşim parametrelerinin tahminleri (Estimations of the binary interaction parameters)

		C	
		1	2
1	B	$\lambda_{11}^{BC} = 0,0436$	$\lambda_{12}^{BC} = -0,0436$
		$\lambda_{21}^{BC} = 0,3013$	$\lambda_{22}^{BC} = -0,3013$
		$\lambda_{31}^{BC} = -0,3449$	$\lambda_{32}^{BC} = 0,3449$

ve ABC üçlü etkileşim terimine ait parametre tahmin değerleri Tablo 16.5'da

Tablo 16.5 Üçlü etkileşim parametrelerinin tahminleri (Estimations of the triple interaction parameter)

		C	
A	B	1	2
1	1	$\lambda_{111}^{ABC} = 0,0406$	$\lambda_{112}^{ABC} = -0,0406$
	2	$\lambda_{121}^{ABC} = 0,1273$	$\lambda_{122}^{ABC} = -0,1273$
	3	$\lambda_{131}^{ABC} = -0,1679$	$\lambda_{132}^{ABC} = 0,1679$
2	1	$\lambda_{211}^{ABC} = 0,0056$	$\lambda_{212}^{ABC} = -0,0056$
	2	$\lambda_{221}^{ABC} = -0,3137$	$\lambda_{222}^{ABC} = 0,3137$
	3	$\lambda_{231}^{ABC} = 0,3081$	$\lambda_{232}^{ABC} = -0,3081$
3	1	$\lambda_{311}^{ABC} = -0,0462$	$\lambda_{312}^{ABC} = 0,0462$
	2	$\lambda_{321}^{ABC} = 0,1865$	$\lambda_{322}^{ABC} = -0,1865$
	3	$\lambda_{331}^{ABC} = -0,1403$	$\lambda_{332}^{ABC} = 0,1403$

verilmiştir. Bu tablolarda yer alan parametrelerin üstündeki indis değişkenleri ve alt indis ilgili değişkenlerin düzeylerini sırasıyla göstermektedir. Bir parametre için tahmin edici ve bu tahmin edicinin varyansı elimizde varsa bu parametre için aynı zamanda hipotez testleri yapılabilir. Yukarıda (4) eşitlikleri ile verilen formüllerle parametrelerin tahmin edicilerinin asimptotik varyansları hesaplandığında (4) eşitliği kullanılarak FTND2 değişkeninin 2. Düzeyi olan orta bağımlılık parametresi için varyansın tahmin değeri

$$Var(\hat{\lambda}_2^A) = \frac{1}{(18)^2} 4 \left( \frac{1}{33} + \frac{1}{20} + \frac{1}{23} + \frac{1}{7} + \frac{1}{4} + \frac{1}{8} \right) + \left( \frac{1}{57} + \frac{1}{12} + \dots + \frac{1}{6} \right) = 0,01135$$

olarak hesaplanır. (5) eşitliği ile verilen test istatistiğinin hesaplanan değeri  $Z(\hat{\lambda}_2^A) = \frac{0,3091}{0,106535} = 2,9014$ ,

$z_{tablo}(0,05) = 1,96$  değerinden büyük olduğu için

$H_0 : \lambda_2^A = 0$  hipotezi red edilir. Yani log-lineer modelde yer alması gereken bir parametredir. FTND2 değişkeninin 2. düzeyinin modele önemli bir etkisi vardır. Yani sigara bağımlılığının orta düzeyde olması hücre frekansı üzerinde önemli bir katkısı vardır. Benzer şekilde ebeveynlerin sigara içme durumu (B) ve Cinsiyet (C) değişkenlerine ait model parametrelerine ait  $z_{hesap}$  değerleri sırasıyla

$$Z(\hat{\lambda}_1^B) = \frac{0,5453}{0,0982} = 5,5530$$

$$Z(\hat{\lambda}_1^C) = \frac{0,4604}{0,0798} = 5,7694 \text{ olup } 1,96 \text{ değerinden büyük}$$

olduğu için için  $H_0 : \lambda_1^B = 0$  ve  $H_0 : \lambda_1^C = 0$  hipotezleri red edilir. Yani log-lineer modelde yer alması gereken parametrelerdir. Ebeveynlerin sigara içme durumu (B) ve Cinsiyet (C) değişkenlerinin 1. düzeylerine ait model parametrelerinin modele önemli bir etkisi vardır. Yani ebeveynlerden birinin sigara içmesi ile cinsiyetin erkek olması hücre frekansı üzerinde önemli bir katkısı vardır.

$\lambda^{ABC}$  parametresinin anlamlılığını test etmek için (5) eşitliği ile verilen formül kullanılarak hesaplanan  $z_{hesap}$  değeri  $Z(\hat{\lambda}_{221}^{ABC}) = \frac{-0,314}{0,146} = -2,1507$  değeri

$-1,96$  değerinden küçük olduğu için s hipotezi red edilir. Yani log-lineer modelde yer alması gereken bir parametredir. Yani, 221 gözesindeki frekansa FTND2 değişkeninin 2. Düzeyi, Ebeveynlerin sigara içme durumu değişkeninin 2. Düzeyi, Cinsiyetin 1. düzeyi etki etmektedir. Bu da [5] iddiasını destekleyen bir sonuçtur. Klasik kontenjans tablo analizlerinden elde edilemeyen bir sonuçtur. Böyle bir analiz SPSS paket programı yardımıyla

#### HILOGLINEAR

B(0 2) C(1 2) A(1 3) /METHOD=BACKWARD  
/CRITERIA MAXSTEPS(10) P(.05) ITERATION(20)  
DELTA(0)  
/PRINT=FREQ RESID ASSOCIATION ESTIM  
/DESIGN .

kodu ile de yapılabilir. Ancak aşağıdaki tabloda da görülebileceği gibi hangi parametrenin hangi düzeyi için anlamlı bir sonuç elde edileceği net bir şekilde görülememektedir.

Tablo 17. Üç yönlü log-lineer analizin SPSS çıktısı (SPSS output of the three-way log-linear analysis)

Etki	Parametre	Tahmin	Std. Hata	Zhesap	p-değeri
BCA	1	,041	,134	,304	,761
	2	,006	,132	,042	,966
	3	,127	,155	,823	,410
BC	4	-,314	,146	-2,155	,031
	1	,044	,098	,444	,657
	2	,301	,113	2,666	,008
BA	1	,194	,134	1,455	,146
	2	-,069	,132	-,524	,600
	3	-,031	,155	-,202	,840
CA	4	,167	,146	1,146	,252
	1	-,294	,111	-2,651	,008

	2	,269	,107	2,530	,011
B	1	,545	,098	5,551	,000
	2	,065	,113	,579	,563
C	1	,460	,080	5,766	,000
A	1	,028	,111	,251	,802
	2	,309	,107	2,902	,004

Tablo 17’de elde edilen SPSS çıktısında her bir ana etki için sahip olduğu düzey sayısının 1 eksiği kadar, etkileşim terimlerinde ise ilgili değişken düzeylerinin 1 eksiklerinin çarpım sayısı kadar parametre tahmin değerlerinin olduğu görülmektedir. Bunun nedeni ise parametre toplamının sifıra eşit olması kısıtının varlığıdır. Buna göre (2) eşitliği ile verilen kısıt gereği her bir değişken için eksik olan parametre tahmini değişkenin diğer düzeylerine ait tahmin değerlerinin toplamının ters işaretlisi olarak bulunabilmektedir.  $p$  değerlerinin yer aldığı sütunda, koyu renk ile gösterilen değerler 0,05 anlam düzeyi ile karşılaştırıldığında  $p < 0,05$  ise oluşturulan hipotez red edilir. Yani bu terim önemlidir ve modelde yer almalıdır. Bu analizin SPSS çıktısında olabilirlik oran test değeri 7,667 ve  $p = 0,467$  olarak bulunduğundan kurulan log- lineer modelimiz yine anlamlı bulunmuştur.

#### 4. SONUÇLAR VE TARTIŞMA (RESULTS AND DISCUSSION)

Bu çalışmada, iki ve üç yönlü kontenjans tabloların log-lineer analizi ile değerlendirilmesinin bilinen klasik kontenjans tablo analizlerine göre üstünlükleri incelenmiştir. Log-lineer modeller yardımıyla daha çok nitel değişken arasındaki etkileşimlerin araştırılması uygulamalı olarak verilmiştir. Uygulamada günümüzün önemli bir bağımlılık problemi olan nikotin bağımlılık düzeyi ile cinsiyet ve ebeveynlerin sigara içme durumları arasındaki ilişki gerçek verilerle araştırılmıştır. Bu çalışmada kullanılan veriler sadece Kırıkkale ilinden toplandığı için takip eden çalışmalarımızda farklı desenlerdeki toplumlardan alınacak verilerle kurulan modellerin daha da anlamlı olacağı düşünülmektedir.

Tablo 6 ile verilen sigara bağımlılık düzeyi ve ebeveynlerin sigara içme durumu değişkenleri için iki yönlü kontenjans tablosu analizi sonucunda kişinin sigara bağımlılık düzeyi değişkeni ile ebeveynlerinin sigara içme durumu değişkenleri birbirinden bağımsız değişkenlerdir, bu iki değişken arasında bir ilişki yoktur sonucu elde edilmiştir. Literatürde yer alan [2-3] sonuçları ile zıt bir sonuç elde edilmesi şüphe uyandırmış ve aşağıda ki daha ileri analizlerin yapılmasına gerek görülmüştür

Tablo 11’deki 2-5 numaralı modeller, kişinin sigara içme düzeyi ile ebeveynlerinin sigara kullanma durumları

arasındaki ilişkinin anlamlı olduğunu gösterirken, Model 1 anlamlı çıkmamıştır. Dahası, ayrı ayrı kurduğumuz Model 3 ve Model 4’ün analizlerinde kurduğumuz hipotez testlerinden gelen 1. tip hata (testin anlam düzeyi)  $2\alpha$  kadar oldu. Bu problemten kaçmak için sabit terim, FTND etkisi, ebeveynin sigara içme durumunun etkisi ve bu değişkenlerin etkileşimlerinin yer aldığı model kurulumu Tablo 11’de Model 5 olarak ifade edilmiştir.

Model 5, sabit terim, kişni sigara bağımlılık düzeyi, ebeveyninin sigara içme durumları ve her iki değişkenin etkileşimlerinin yer aldığı doymuş log-lineer modeldir. Analiz sonucunda bağımlılığın az olduğu düzeye ait ana etki -0,0554 iken ebeveynin her ikisinin de sigara içmediği düzeye ait ana etki 0,5164 olarak hesaplanmıştır. Ayrıca az bağımlı ve ebeveynin sigara içmediği düzey için etkileşim değeri 0,1685 olup çok bağımlı ve ebeveynin ikisinin de sigara içtiği düzey için etkileşim değeri 0,1605 olarak hesaplanmıştır.

Tablo 14’de yer alan parametrelerin önemlilik testi sonuçlarına göre kişinin sigara bağımlılık düzeyi değişkeni ve ebeveynlerinin sigara içme durumu değişkenleri arasında etkileşim olmadığı ortaya çıktı. Bu durum [2-3] çalışmalarında verdiği sonuçla çeliştiği için modele üçüncü bir değişken olan cinsiyet değişkeni de eklenerek daha ayrıntılı bir analiz olarak 3 yönlü kontenjans tablosunda log-lineer analiz uygulanmıştır. Cinsiyet, ebeveynlerin sigara içme durumu ve nikotin bağımlılık düzeyi değişkenleri ile kurulan bu 3 yönlü kontenjans tablosuna log-lineer analiz uygulandığında sigara bağımlılığının orta düzeyde olması hücre frekansı üzerinde önemli bir katkısının olduğu görülmüştür. Ebeveynlerden birinin sigara içmesi ile cinsiyetin erkek olmasının model frekansı üzerinde önemli bir katkısının olduğu görüldü.

Bu çalışmadaki sonuçlarımıza göre de Kırıkkale’de yaşayan ve sigara kullanan bir bireyin, sigara bağımlılık düzeyinin düzeyinin orta, cinsiyetinin erkek ve ebeveynlerinin birinin sigara içmesi log-lineer modeldeki frekansı artıran çok önemli bir etkidir.

Bu verileri toplamaktaki amacımız sigara içmeyen bir toplum olabilmemiz için öneriler geliştirmektir. Buna göre kadınlara nazaran sosyal çevrede daha çok yer alan bir erkek birey olmak ve ebeveynlerinin birinin sigara içiyor olması bağımlılık üzerinde çevresel ve genetik etkilerin beraber araştırılması gerektiğini ortaya koymaktadır. Sigara bağımlılığını önleme çalışmalarında, ebeveynlerin kendi çocuklarının gelecekte sağlıklı bireyler olması için, öncelikle anne ve babanın sigara konusunda daha da bilinçlenmesi önemlidir. Bilimsel olarak zararı ispatlanmış olan sigaranın kullanılmaması gerektiğini sözel olarak

çocuklarımıza ifade etmenin tek başına yeterli olmayıp, aynı zamanda ebeveynlerin kendi davranışlarıyla da iyi bir rol model olması gerekmektedir. Türk toplumunda erkeklere yüklenen rol model gereği erkeklerin sigara içmedeki özentilerinin önüne geçilmesine yönelik çalışmalar yapılmalıdır.

### 5. TEŞEKKÜR (ACKNOWLEDGEMENT)

Çalışmamıza anket formunu doldurarak katılan herkese minnettarız. Bu çalışmada, bize veri toplama hususunda gönüllü olarak yardımcı olan Kırıkkale Üniversitesi İstatistik Bölümü öğrencilerine teşekkürlerimizi sunarız.

### KAYNAKLAR (REFERENCES)

- [1] L.İ. Yargıç ve H. Baykan, "Severe mental disorders, depression and smoking cessation", *Anadolu Psikiyatri Dergisi*, cilt 14, no. 1, pp. 77-83, 2013.
- [2] E. Abay ve İ. Ateş, "Bağımlılığın genetiği", *Bağımlılık Dergisi*, cilt 2, no. 2, pp. 68-70, 2001.
- [3] P. F. Sullivan, K. S. Kendler. "The genetic epidemiology of smoking", *Nicotine & Tobacco Research*, cilt 1, no. 2, pp. 51-57, 1999.
- [4] A. Yorgancıoğlu ve A. Esen, "Sigara Bağımlılığı ve Hekimler", *Türk Toraks Dergisi*, cilt 1, no. 1, pp. 90-95, 2000.
- [5] O. Okutan ve T. Çalışkan, "Sigara ile ilişkili İnterstisyel Akciğer Hastalıkları", *Solunum Dergisi*, cilt 13, no. 3, pp. 131-139, 2011.
- [6] "World Health Organization (WHO). WHO report on the global tobacco epidemic", 2011a. [Çevrimiçi]. Available: [http://www.who.int/tobacco/surveillance/policy/country\\_profile/](http://www.who.int/tobacco/surveillance/policy/country_profile/). [Erişildi: 1 Mayıs 2014].
- [7] "World Health Organization (WHO). WHO global youth tobacco survey", 2011b. [Çevrimiçi]. Available: <http://www.who.int/tobacco/surveillance/guys/en/>. [Erişildi: 1 Mayıs 2014].
- [8] J. B. Bricker, A. V. Peterson, M. R. Andersen, B. G. Leroux, K. B. Rajan ve I. G. Sarason, "Close friends', parents', and older siblings' smoking: reevaluating their influence on children's smoking", *Nicotine & Tobacco Research*, cilt 8, no. 2, pp. 217-226., 2006.
- [9] G.F. El-Sharkawy, "Cigarette smoking among University students: family related and personal risk factors", *Journal of American Science*, cilt 7, no. 3, pp. 1-9, 2011.
- [10] S. Öncel ve F. Aliev, "The Influence of Family Factors on Smoking Behavior in Turkey", *International Journal of Statistics in Medical Research*, cilt 3, no. 2, pp. 116-125, 2014.
- [11] S. Yurt Öncel, D. M. Dick, H. H. Maes ve F. Aliev, "Risk Factors Influencing Smoking Behavior: A Turkish Twin Study", *Twin Research and Human Genetics*, cilt 17, no. 6, pp. 563-573, 2014.
- [12] S. Erbaş ve S. İçöz, "Sosyal Bilimlerde Niteliksel Veri Analizi Üzerine Bir Çalışma", *Ekonomik Yaklaşım*, cilt 9, no. 9, pp. 91-94, 1998.
- [13] S. Bülbül, "Üç boyutlu çapraz tablolarda logaritmik doğrusal analiz: Çocuk işgücü değişkenleri arasındaki etkileşimler", *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, cilt 25, no. 2, pp. 41-70, 2006.
- [14] D. Altaş, M. Sağırılı ve S. Giray, "1. Altaş, D., Sağırılı M., Giray S., 2006 Yurtdışında Çalışıp Türkiye'ye dönen akademisyenlerin eğitim durumları, gidiş ve dönüş sebepleri arasındaki ilişki yapısının log-lineer modeller ile incelenmesi", *Marmara Üniversitesi İ.İ.B.F. Dergisi*, cilt XXI, no. 1, pp. 401-421, 2006.
- [15] Z. Filiz, "Üç yönlü log-lineer modeller ile üniversite öğrencilerinin sigara, alkol ve nargile içme nedenlerini etkileyen faktörlerin belirlenmesi", *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, cilt 8, no. 2, pp. 225-250, 2007.
- [16] E. Şıklar, V. Yılmaz, D. Çoşkun "Eskişehir'deki Üniversitelerde Görevli Akademik Personelin İş Tatmini ve Duygusal Tükenmişliklerinin Log-Linear Modeller ve Correspondence Analizi ile İncelenmesi", *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, cilt 26, no. 2, pp. 113-134, 2011.
- [17] S. Mete ve A. Ünsal, "Kategorik veriler için logaritmik doğrusal modeller ve göç istatistikleri üzerine bir uygulama", *Aksaray Üniversitesi İİBF Dergisi*, cilt 2, no. 1, pp. 9-20, 2010.
- [18] V. Yılmaz ve C. Aktaş, "Üç boyutlu kontenjans tablolarının analizinde log-linear modellerin kullanımı ve trafik kazalarına uygulanması", *Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, no. 2, pp. 169-182, 2001.
- [19] M. Kateri, *Contingency table analysis: Methods and implementation using R*, 2010.
- [20] B. Lawal, *Categorical Data Analysis with SAS and SPSS Applications*, London: Lawrence Erlbaum Associates, 2003.
- [21] R. Christensen, *Log-Linear Models and Logistic Regression*, New York : Springer-Verlag, 1997.

- [22] D.C. Howell, *Statistical Methods for Psychology (Seventh Edition)*, Belmont: Wadsworth Cengage Learning, 2009.
- [23] S. K. Lee, “On the asymptotic variances of u terms in loglinear models of multidimensional contingency tables”, *Journal of the American Statistical Association*, cilt 72, no. 358, pp. 412-419, 1977.
- [24] K.O. Fagerström, “Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment”, *Addictive Behaviors*, cilt 3, pp. 235-241, 1978.

