Atatürk Ü.Zir.Fak.Der. 24 (2), 21- 35, 1993.

ÜÇ - BOYUTLU ÇAPRAZ - SINIFLANDIRILMIŞ KATEGORİK VERİLERE LOG-LİNEAR MODELLERİN UYDURULMASI

Hüdaverdi BİRCAN (1)
Necati YILDIZ (1)

ÖZET : *Biyoloji, tıp ve sosyal bilim dallarındaki araştırmalarda ölçülen karakterlerin bazen kalitatif (nicel) olması nedeniyle, elde edilen kesikli değişkenlerin analizinde bilinen parametrik metotlar elverişli olmamaktadır. Bu özellikteki veriler çapraz-sınıflanmış tablolar şeklinde takdim edilir. Çapraz sınıflandırılmış kategorik verilerin iki-boyutlu olması durumunda Pearson'un χ^2 istatistiği ve G^2 olabilirlik-oran istatistiği kullanılmaktadır. Bir çok durumda bir sonuca tesir eden faktör sayısı ikiden fazladır. Bu nedenle yaygın olarak kullanılan çapraz-sınıflandırılmış tablolar çok boyutludur. Çok boyutlu tabloların analizi için; 1- çok boyutlu tablodan elde edilecek tüm iki boyutlu marjinal tabloların ayrı ayrı analiz edilmesi, 2- çok boyutlu tablolara log-linear modellerin uydurulması şeklinde iki yol vardır.*

Üç faktörlü çapraz-sınıflandırılmış verilerin analizinde; mümkün olan iki faktörlü marjinal tabloların elde edilerek ayrı ayrı klasik χ^2 testinin uygulanması, tüm ikili ilişkilerin aynı zamanda denenmelerine müsaade etmediği gibi üç faktör interaksiyonunu da ihmal eder. Bundan dolayı bu çalışmada, log-linear yaklaşımının özellikleri ve uygulanması üzerinde durulmuştur.

FITTING THE LOG-LINEAR MODELS TO THREE - DIMENSIONAL CROSS-CLASSIFIED CATEGORICAL DATA

SUMMARY : *Some parametric methods of analysis are not convenient for non-normal-discrete data obtained from social, biological and medical research. This type of data generally are presented in the cross-classified categorical table. Well know Pearson's χ^2 and likelihood-ratio G^2 statistics can be used for the analysis of two-way contingency tables. In many cases a researcher deals with more than two factors effect on response variable. For this reason, there are two different procedures*

(1) Atatürk Üniversitesi Ziraat Fakültesi Zootekni Bölümü, 25240 Erzurum.

for the analysis of multidimensional tables : 1- Separate analysis of all possible two by two tables obtained from multidimensional tables. 2- Fitting the log-linear models directly to multidimensional tables.

The analysis of three-dimensional categorical tables by sperating all possible two by two marjinal tables is not satisfactory. Although such approch sometimes gives limited insight about the relationships among the variables, does not allow for the simultaneous examination of these pairwise relationships and ignores the possibility of three-factor interactions among the variables. Therefore the features and applications of log-linear models fitting approach were discussed in the detail on this study.

GİRİŞ

Biyolojik, tıp ve sosyal bilim dallarındaki araştırmalardan elde edilen veriler genellikle sayımla belirlenir. Bu sahalardaki araştırıcıların ölçtükleri karakterlerin kalitatif (nicel) olması, uygun değişken tipinin kesikli olmasını gerektirir. Böyle verilerin analizinde bilinen parametrik metodlar elverişli olmamaktadır. Bu özellikteki verilerin analizinde kullanılan en yaygın metod ihtimal tabloları olarak belirtilen çapraz-sınıflanmış tabloların analizidir.

Çapraz-sınıflandırılmış kesikli ve çok değişkenli verilerin analizi, Pearson (1900) ve Fisher (1922) zamanından günümüze kadar literatürde önemli bir yer işgal etmiştir. Sayımla tesbit edilen çapraz-sınıflandırılmış kategorik verilerin iki-boyutlu olması halinde, Pearson'un χ^2 istatistiği ve ilk defa Wilks (1935) tarafından teklif edilen G^2 olabilirlik oran istatistiği kullanılarak analiz edilebilmektedir. Halbuki birçok durumlarda bir sonuca tesir eden faktör sayısı ikiden fazla olmaktadır. Yaygın olarak karşımıza çıkan bu çok-boyutlu (çok-faktörlü) tabloların analizi için genellikle iki standart yol izlenmiştir :

1) Çok boyutlu tablodan, mümkün olan iki-boyutlu marjinal toplamları elde ederek ayrı ayrı analiz etmek. Bu metodlar değişkenler arasındaki ilişki hakkında belirli bir doğrulukta bilgi sağlamış olmakla birlikte, aşağıdaki mahzurları da taşımaktadır :

a) Diğer değişkenler dikkate alındığında, ilişkili kategorik değişken çiftleri arasındaki marjinal ilişkiyi tam ayıramaz.

b) Değişkenler arasındaki üç faktör ve daha yüksek dereceli intiraksiyonların varlığını ihmal eder.

c) Tüm ikili ilişkilerin aynı zamanda denenmelerine müsade etmez.

2) Varyans analizi gibi, bilinen bazı parametrik metodları uygulamak. Bu yolun rahatlıkla uygulanabilmesi varyans analizi farayizelerinden; x_{ij} gözlemlerinin varyansları homojen, gözlemlerin normal dağılışa uygun ve uynı zamanda toplanabilirlik faraziyetlerinin geçerli olması gerekmektedir.

Çapraz-sınıflandırılmış bir ihtimal tablosundaki veriler; multinomial, binom ve poisson gibi dağılışlara uygun iseler, bu faraziyelerden birçoğu yerine getirilemez. Böyle bir durumda faraziyeleri geçerli kılmak için akla bazı transformasyonlar gelmektedir. Ancak bu transformasyonlarında yukarıdaki şartların tümünü birden sağlaması nadirdir (Fienberg, 1970a).

Gerek çok-boyutlu tabloların tüm iki-boyutlu marjinal toplamlarının ayrı ayrı analizlerinin ve gerekse varyans analizinin uyarıda bahsedilen mahzurlarından dolayı, çok-boyutlu kategorik verilerin analizinde log-linear modellerin uygulanmasına geçilmiştir. Log-linear modeller tablo boyutlarına tekabül eden değişkenler arasındaki yapısal ilişkiyi açıklamak için kullanılır (Fienberg, 1977). Çok-boyutlu tabloların analizi; Bartlett (1935) tarafından 2x2x2 boyutlu tablolarda ikinci dereceden interaksiyonun olmaması durumunu test etme ile ilgili ilk makalesinden günümüze kadar istatistiksel literatürlerde önemli bir yer işgal etmiştir.

Birch (1963), üç yönlü ve çok-yönlü ihtimal tablolarındaki interaksiyonları, beklenen frekansların logaritmalarının linear kombinasyonları olarak açıklamıştır. Araştırıcı üç-yönlü tablolar için, üç değişken arasında yapısal ilişkiye dayanarak açıklanan en az beş loglinear modelin olduğunu göstermiştir. Ayrıca tüm bu modeller için en yüksek olabilirlik tahminleri (maximum likelihood estimation) nin çözümünü vermiştir.

Goodman (1964), m-boyutlu $d_1 x d_2 x ..... x d_m$ ihtimal tablosunda r. dereceden interaksiyonların (r = 0, 1, 2, ..., m-1) bir tarifini vermiştir. Bu interaksiyonların belirlenmiş bir alt setinin sıfıra eşit olduğu hipotezini test etmek için metodlar sunmuştur. Goodman (1964)'nın sunduğu metodlar, üç-boyutlu bir tablodaki sıfır hipotezini test etmek için Goodman (1963) ve Plackett (1962) tarfından ilk olarak teklif edilen metodların genelleştirilmiş şeklidir. Good (1956), üç-boyutlu bir tabloda sıfır hipotezini test etmek için bir test teklif etmiştir. Bu metod ilk olarak Bartlett (1935), Roy ve Kastenbaum (1956) ve Darroch (1962) tarafından teklif edilen testlerin daha genelleştirilmiş şeklinden ibarettir.

Feinberg (1970b), çok-boyutlu ihtimal tablolarının analizi için, varyans analizinde kullanılan modellere benzer modeller sunmuştur. Farklı modellerin beklenen hücre değerlerinin hesaplanması için, bu modellere ait ençok olabilirlik tahminlerini vermiştir. Fienberg (1970a), belli bir modelde uyumun iyiliğini kontrol etmek için, iki farklı test istatistiğinin kullanılabileceğini açıklamıştır. Bunlardan birincisi, Pearson'un χ^2 istatistiği, ikincisi ise; G^2 olabilirlik oran istatistiğidir.

Bishop ve ark. (1975), logaritmik skalada linear olan modellerin, ihtimal tablolarının yapısı ile tarif edileceğini göstermişlerdir. İhtimal tabloların analizinde loglinear modellerin uygulanmasının geniş bir tartışmasını vermişlerdir. Hiyerarşik (iç-içe) modellere ait beklenen hücre değerlerinin en yüksek olabilirlik tahminlerini açıklamışlardır. Ayrıca üç ve dört-boyutlu tablolar için beklenen değerlerin doğrudan tahminlerini göstermiştirler. Uygun modelin seçiminde, uyumun-iyiliği istatistiklerinden Perason'un χ^2 ve olabilirlik-oran istatistiği G^2'yi kullanmışlardır.

Fienberg (1977), çapraz-sınıflandırılmış verilerin analizi için log-linear modellerin geniş bir açıklamasını sunmuştur. Log-linear modellerin oluşturulmasındaki hiyerarşik esasları, bu modeller altında beklenen değerlerin en yüksek olabilirlik tahminlerini ve "ikinci-dereceden interaksiyonsuz" modele ait beklenen değerlerin tekrarlı tahmini için bir yöntem vermiştir. Aynı zamanda bu yöntemin diğer modellere ait beklenen değerlerin hesaplanmasında da kullanılabileceğini göstermiştir. Modellerin uyumu için, iki istatistik kullanmıştır; birincisi Pearson'un χ^2 istatistiği ve ikincisi olabilirlik-oran G^2 istatistiğidir.

Khamis ve Hinkelmann (1984), hastalık-genotip ilişkisi problemlerinin daha fazla iç yüzünü anlamayı sağlamak için log-linear model analizinin, kullanılabileceğini göstermişlerdir. Cressie ve Read (1989), Pearson'un χ^2 ve olabilirlik-oran G^2 istatistiklerine 40 yıllık literatür çerçevesinde karşılaştırmalı olarak incelemişlerdir.

Log-linear modellerin çapraz-sınıflandırılmışı kategorik verilere (üç-boyutlu) uygulanması ile; tıp, biyolojik ve sosyal bilim dallarındaki araştırıcıların ihtiyaç duyduğu bilgileri temin edecek tarzda analiz yapılabilmektedir. Bu çalışmada, üç-faktörlü çapraz-sınıflandırılmış kategorik verilerin analizinde klasik χ^2 testleri ile, log-linear maddelerin uydurulması metodlarının uygulaması, özellikleri ve sonuçları karşılaştırmalı olarak tartışılacaktır.

MATERYAL VE METOT

Materyal

Bu araştırmada, materyal olarak Kılıç (1984) tarafından elde edilen verilerden faydalanılmıştır. Kılıç (1984) tarafından yapılan (Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı) çalışmadan elde edilen veriler; üç-boyutlu çapraz sınıflandırılmış verilerin en tipiği olan 2x2x2 boyutlu tabloların değerlendirilmesinde kullanılmıştır. Bu araştırma; 42 behçet hastası olan bir grup ile, 38 kişilik kontrol grubunu içermektedir. Söz konusu hasta ve kontrol grubu için sakroiliitis'in görülme sıklığı araştırılmıştır.

Metot

Üç boyutlu tablolarda; x_{ijk}, i. sıra, j. sutun ve k. dizeydeki (3. faktör) gözlemleri; P_{ijk}, (i, j, k) hücrelerine düşen bireylerin ihtimalini ve M_{ijk} ise (i, j, k) hücrelerine düşen beklenen değerleri göstermektedir. Burada, i = 1, ... I, j = 1, ... J ve k = 1, ..., K 'dır. N, toplam müşehade sayısıdır. "+" indisi, bir ya da daha çok indisleri gösteren ortak değerli tüm hücrelerin karşılıklı toplanacağını ifade etmektedir. Mesela,

$$M_{+jk} = \sum_{i=1}^{I} M_{ijk},$$

j ve k'yi temsil eden ortak değerli tüm hücrelerin toplamıdır.

$$x_{ij+} = x_{ij1} + x_{ij2}, \quad \ldots\ldots\ldots\ldots \quad k = 1, 2$$

$$x_{j++} = x_{i11} + x_{i12} + x_{i21} + x_{i22}$$

$$= x_{i1+} + x_{i2+} \quad \ldots\ldots\ldots\ldots \quad j = k = 1, 2$$

ve

$$x_{+++} = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} x_{ijk} = \sum_{i=1}^{2} x_{i++}$$

şeklindedir. Şayet tablonun boyutlarına tekabül eden 3 değişken bağımsız ise;

P_{ijk} = P (1. değişkenin kategorisi = i) * P (2 değişkenin kategorisi = j)

*P (3. değişkenin kategorisi = k)

dir. (i, j, k) hücresine ait beklenen değerlerin tahminini elde etmek için,

$$\hat{M}_{ijk} = \frac{x_{i++}}{N} \cdot \frac{x_{+j+}}{N} \cdot \frac{x_{++k}}{N} \cdot N \quad \ldots\ldots\ldots\ldots \quad 2.1$$

eşitliği kullanılır. Bu eşitliğin tabi logaritması alınırsa,

$$\log M_{ijk} = \log x_{i++} + \log x_{+j+} + \log x_{++k} - 2 \log N \quad \ldots\ldots\ldots \quad 2.2$$

eşitliği bulunur. Tahminlenmiş beklenen değerlerin logaritmalarının toplanabilir şekli, varyans analizi notasyonunu hatırlatmaktır. Tahminlenmiş M_{ijk} değerleri yerine M_{ijk} parametreleri dikkate alınarak aşağıdaki eşitlikler yazılabilir (2x2x2 tabloda);

$$U = \frac{1}{IJK} \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} \log M_{ijk} \quad \ldots\ldots \quad 2.3$$

$$= -2\log N + \frac{1}{2} \left[\sum_{i=1}^{2} \log M_{i++} + \sum_{j=1}^{2} \log M_{+j+} + \sum_{k=1}^{2} \log M_{++k} \right]$$

U, beklenen sayıların logaritmalarının ortalamasıdır.

$$U_{1(i)} = \frac{1}{JK} \sum_{j=1}^{2} \sum_{k=1}^{2} \log M_{ijk} - U, \quad \ldots\ldots \quad 2.4$$

Benzer şekilde,

$$U_{2(j)} = \frac{1}{IK} \sum_{i=1}^{2} \sum_{k=1}^{2} \log M_{ijk} - U, \quad \ldots\ldots \quad 2.5$$

Buna göre $U_{3(k)}$ da, $U_{1(i)}$ ve $U_{2(j)}$ şeklinde yazılabilir. $U_{1(i)}$, $U_{(2)}$ ve $U_{3(k)}$, genel ortalama U'dan sapmaları gösterir. log M_{ijk}, varyans analizi tipi bir notasyonda yazılabilir.

$$\log M_{ijk} = U + U_{1(i)} + U_{2(j)} + U_{3(k)} \quad \ldots\ldots \quad 2.6$$

Burada;

$$\sum_{i=1}^{2} U_{1(i)} = \sum_{j=1}^{2} U_{2(j)} = \sum_{k=1}^{2} U_{3(k)} = 0 \quad \ldots\ldots \quad 2.7$$

kısıtlamaları geçerlidir.

Üç değişkenin bağımsız olmadığı farz edilirse, aşağıdaki gibi üç değişken arasındaki yapısal ilişkiye dayanarak açıklanabilen 4'den fazla model tipi vardır (Birch, 1963; Fierberg, 1970a; Bishop ve ark., 1975; Fierberg, 1977).

1) 3. değişken, 1. ve 2. değişkenle birlikte bağımsızdır (Bu modelin diğer iki şekli vardır). $U_{13} = U_{23} = 0$

2) Sadece bir ikili ilişki (U_{12}) bağımsızdır. (Bu modelin de diğer iki şekli vardır) $U_{12} = U_{123} = 0$

3) Üçüncü değişkenin değeri tarafından etkilenmemiş, herbir iki değişkenli, üç esas değişken arasındaki ikili ilişki $U_{123} = 0$

4) Tüm üç değişkenle ilişkili olan, ikinci dereceli interaksiyon (üç faktör etkili). Yukarıda listelenen tüm modeller için genel loglinear model,

$$\log M_{ijk} = U + U_{1(i)} + U_{2(j)} + U_{3(k)} + U_{12(ij)} + U_{13(ik)} + U_{23(jk)} + U_{123(ijk)} \quad \ldots \quad 2.8$$

dir. Burada genellikle varyans analizindeki gibi,

$$\sum_i U_{1(i)} = \sum_j U_{2(j)} = \sum_k U_{3(k)} = 0$$

$$\sum_i U_{12(ij)} = \sum_j U_{12(ij)} = \sum_i U_{13(ik)} = \sum_k U_{13(ik)} = \sum_j U_{23(jk)} =$$

$$= \sum_k U_{23(jk)} = 0 \quad \text{..............................} \quad 2.9$$

$$\sum_i U_{123(ijk)} = \sum_j U_{123(ijk)} = \sum_k U_{123(ijk)} = 0$$

kısıtlamaları vardır.

Örnekleme modelleri; poisson, multinomial ve çarpımlı multinomial (product multinomial) olduğu zaman log-linear modellerin tümü için beklenen hücre değerlerinin en yüksek olabilirlik tahminlerinin aynı olduğunu Haberman (1974) göstermiştir.

Eşitlik (2.1), tamamen bağımsız model için tahminlenmiş beklenen değerleri vermektedir. Diğer modeller için de tahminlenmiş beklenen değerler Birch (1963) ve Fienberg (1977) tarafından verilmiştir. Buna göre eşitlik (2.7)'deki genel log-linear model ele alınırsa,

$$U_{12(ij)} = U_{123(ijk)} = 0 \quad \text{..} \quad 2.10$$

modeli için tahminlenmiş beklenen değerler,

$$\hat{M}_{ijk} = \frac{X_{i+k}\, X_{+jk}}{X_{++k}} \quad \text{..} \quad 2.11$$

formülünden;

$$U_{13(ik)} = U_{23(jk)} = U_{123(ijk)} = 0 \quad \text{..} \quad 2.12$$

modeline ait tahminlenmiş beklenen değerler ise,

$$M_{ijk} = X_{ij+}\; X_{++k}/N \quad \text{..} \quad 2.13$$

formülünden elde edilir. Son olarak genellikle "ikinci-dereceden interaksiyonsuz" model olarak isimlendirilen, model 53) gözönüne alınabilir.

$$U_{123(ijk)} = 0 \quad \text{..} \quad 2.14$$

tüm i, j, k için. Bu model için tahminlenmiş beklenen değerler doğrudan elde edilemez. Bundan dolayı bu model için tahminlenmiş beklenen değerler Bishop (1967) ve Fienberg (1970b) tarafından verilen tekrarlı yöntem yoluyla bulunabilmektedir. Tahminlenmiş beklenen değerlerin tekrarlı yöntem ile hesaplanmasında;

1. Adım : Tablonun her hücresindeki ilk değer 1'e eşit olacaktır. Yani,

$M_{ijk}^{(0)} = 1,$ tüm i, j, k için

sonra, v = 0 için diğer adımlara geçilir.

2. Adım : 1. adımdaki girişler ($M_{ijk}^{(0)}$); X_{ij+} gözlenen marjinal toplamların 1. adımdaki beklenen marjinal toplamlara ($M_{ij+}^{(0)}$) oranı ile çarpılır. Yani,

$$\hat{M}_{ijk}^{(3v+1)} = \hat{M}_{ijk}^{(3v)} \left(\frac{X_{ij+}}{\hat{M}_{ij}^{(3v)}}\right) \qquad 2.15$$

dir.

3. Adım : 2.adımdaki girişler ($M_{ijk}^{(1)}$); X_{i+k} gözlenen marjinal toplamların 2. adımdaki beklenen marjinal toplamlara ($M_{i+k}^{(1)}$) oranı ile çarpılır.

$$\hat{M}_{ijk}(3v+2) = \hat{M}_{ijk}(3v+1)\left(\frac{X_{i+k}}{\hat{M}_{i+k}(3v+1)}\right) \qquad 2.16$$

4. Adım : 2. ve 3. adıma benzer olarak; 3. adımdan girişler ($M_{ijk}^{(2)}$); X_{+jk} gözlenen marjinal toplamların, 3. adımdaki beklenen marjinal toplamlara ($M_{+jk}^{(2)}$) oranına çarpılması ile elde edilir.

$$\hat{M}_{ijk}^{(3v+3)} = \hat{M}_{ijk}^{(3v+2)}\left(\frac{X_{+jk}}{\hat{M}_{+jk}^{(3v+2)}}\right) \qquad 2.17$$

Bu durum tekrarın ilk dizisini tamamlar. Tekrar işlemi v = 1, 2, ..., n için (2.15) - (2.17)'deki eşitliklere göre devam ettirilerek tahminlenmiş beklenen değer (M_{ijk}); bir sonraki işlemde bir öncekine göre fazla değişmiyorsa tekrar işlemi durdurulur.

Tahminlenmiş beklenen hücre değerleri log-linear modeller için yukarıda açıklanan direk ve tekrarlı tahmin metotlarından elde edilmektedir. Bu söz konusu modellerin uyumlarının testinde, Pearson'un χ^2 ve olabilirlik-oran G^2 istatistiği kullanılır. Pearson'un (1) χ^2 istatistiği;

$$X^2 = \sum \frac{(\text{Gözlenen-Beklenen})^2}{\text{Beklenen}} \qquad 2.18$$

olup, Wilks (1935) tarafından teklif edilen olabilirlik-oran G^2 istatistiği ise;

$$G^2 = 2\sum (\text{Gözlenen}) \log \left(\frac{\text{Gözlenen}}{\text{Beklenen}}\right) \qquad 2.19$$

şeklindedir. G^2, asimtotik olarak Pearson'un (1900) χ^2 istatistiğine eşittir.

Tahminlenmiş beklenen değerler yardımı ile her model için ayrı ayrı χ^2 ve G^2 istatistikleri hesaplanır. χ^2 ve G^2 istatistikleri aşağıdaki formülle verilen serbestlik dereceli χ^2 dağılışına sahiptir.

Serbestlik Derecesi = Hücre Sayısı - Uydurulan Parametre Sayısı 2.20

Bu formül, 5 farklı modelin herbirisine uygulanırsa Tablo 1'de gösterilen serbestlik dereceleri elde edilir.

Tablo 1'de log-linear modellerin mümkün olan tüm çeşitleri ele alınmamıştır. Mesela gözönüne alınmayan modellerden birisi,

$$\log M_{ijk} = U + U_{1(i)} + U_{2(j)} + U_{3(k)} + U_{123(ijk)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad 2.21$$

Tablo 1. Üç-Boyutlu Tablolar İçin Değişik Loglinear Modellerle İlişkili Olan Serbestlik Dereceleri.

Table 1. Degrees of Freedom Associated With Various Loglinear Models for Three Dimersinal Tables.

Model	Kısaltma	Uydurulan Parametre Sayısı*	Serbestlik Derecesi*
$U+U_1+U_2+U_3$	[1][2][3]	4 [1+(I-1)+(J-1)+(K-1)]	4 (IJK-I-J-K+2)
$U+U_1+U_2+U_3+U_{12}$	[12][3]	5 [1+(I-1)+(J-1)+(K-1)+ (I-1)(J-1)]	3 (K-1)(J-1)
$U+U_1+U_2+U_3+U_{12}+U_{23}$	[12][23]	6 [1+(I-1)+(J-1)+(K-1)+ (I-1)(J-1)+(J-1)(K-1)]	2 [J(I-1)(K-1)]
$U+U_1+U_2+U_3+U_{12}+U_{23}+U_{13}$	[12][13][23]	7 [1+(I-1)+(J-1)+(K-1)+ (I-1)(J-1)+(I-1)(K-1)+ (J-1)(K-1)	1 [(I-1)(J-1)(K-1)]
$U+U_1+U_2+U_3+U_{12}+U_{23}+U_{13}+U_{123}$	[123]	8 IJK	0

* Rakamlar 2x2x2 tablosuna, alttaki formüller ise, IxJxK boyutlu tablolara aittir.

dir. Daha yüksek-dereceden U terimlerinin daha düşük-dereceden U terimlerinden sapmalarının nasıl ölçüldüğü eşitlik (2.4) ve (2.5)'de verilmiştir. Bu yorumu muhafaza etmenin kuralı da, daha yüksek-dereceli terimler yalnızca ilişkili olan daha düşük-dereceli terimleri ihtiva eden hiyerarşik setle modellerin sıralanmasıdır. Böylece U_{123}; U_{12}, U_{13} ve U_{23}'ün tümü bir modelde yeralmadıkça, modelde bulunamaz. Açık olarak (2.21)'de verilen eşitlik bu manada hiyerarşik bir model değildir. Verilere (2.21)'deki gibi hiyerarşik olmayan modellerin uygunluğunu da göz önüne almak mümkündür. Ancak bu durumda, tekrarlı uyum işlemi yoluyla tahminlenmiş beklenen değerler hesaplanamaz. Haberman (1974), hiyerarşik olmayan log-linear modellerin, olabilirlik eşitliklerini elde etmek için doğrudan kullanılabilen bazı genel sonuçlar vermektedir. Hiyerarşik olmayan modellerin önemli dezavantajı, onların yorumlanmasıdır.

Üç-boyutlu tablolar için, verilere herhangi birinin uyabileceği 8 mümkün hiyerarşik model vardır. Pearson'un χ^2 ve olabilirlik-oran G^2 istatistikleri, bu 8 modelden eğer sadece birinde önemsiz ise; ilgili modele ait hipotez kabul edildiğinden bu model seçilir. Eğer birden fazla modelde test istatistikleri önemsiz ise o takdirde uygun modelin seçimi için yardımcı olacak metotlara ihtiyaç vardır.

Üç-boyutlu çapraz sınıflandırılmış verilere log-linear modellerinin uydurulmasında, ilgili tüm modeller için Pearson'un χ^2 ve olabilirlik-oran G^2 istatistiğini veren Statgraphics (1988) (Statistical Graphics System), paket programı kullanılmıştır.

TARTIŞMA VE SONUÇLAR

Tablo 2, Kılıç (1984) tarafından elde edilen verilerin çapraz sınıflandırmasına ait üç-boyutlu tablo ve bununla ilgili tüm iki-boyutlu tabloları göstermektedir. Tablo 2'deki veriler multinomial örnekleme metodu yoluyla meydana getirilmiştir. Sakroiliitis durumu cevap değişkeni, hastalık ve cinsiyet ise açıklayıcı değişkenler olarak düşünülebilir.

Tablo 2 b, c ve d'deki iki-boyutlu tablolara, bağımsızlık şartı altında Pearson'un χ^2 istatistiği uygulanırsa; sırasıyla 9.21, 0.086 ve 6.38 değerleri elde edilir. Bu değerlerden Tablo 2 b'ye ait $X^2 = 9.21$ değeri, 1 serbestlik dereceli χ^2 dağılışında çok önemlidir ($P < 0.01$). Yani, hastalık durumu ile sakroiliitis durumu birbiriyle çok önemli derecede ilişkilidirler. Tablo 2 d'ye ait $X^2 = 6.38$ değeri de 1 serbestlik dereceli χ^2 dağılışında önemlidir ($P<0.05$). Bu durum sakroiliitis durumu ile cinsiyet birbiriyle önemli derecede ilişkili olduğunu göstermektedir. Ancak, Tablo 2 c'ye ait değişkenler birbirinden bağımsızdır. Çünkü $X^2 = 0.086$ değeri, 1 serbestlik

Tablo 2. 80 Ferd İçin; Cinsiyet, Sakroiliitis Durumu ve Hastalık Durumuna Göre Üç-Boyutlu Çapraz-Sınıflandırma.

(a) Gözlenen Değerler.

(a) Observed Data.

Cinsiyet	Erkek		Kadın	
Sak. Durumu / Hastalık	Var	Yok	Var	Yok
Hasta	17	13	7	5
Kontrol	1	25	8	4

(b) Hastalık ve Sakroiliitis Durumu Değişkenlerinin 2x2'lik Marjinal Toplamı.

(b) 2x2 Marjinal Total.

Sak. Durumu / Hastalık	Var	Yok	Toplam
Hasta	24	18	42
Kontrol	9	29	38
Toplam	33	47	80

(c) Hastalık ve Cinsiyete Göre 2x2'lik Marjinal Toplam

(c) 2x2 Marjinal Total

Cinsiyet / Hastalık	Erkek	Kadın	Toplam
Hasta	30	12	42
Kontrol	26	12	38
Toplam	56	24	80

(d) Cinsiyet ve Sakroiliitis Durumuna göre 2x2'lik Marjinal Toplam.

(d) 2x2 Marjinal Total.

Sak. Durumu / Cinsiyet	Var	Yok	Toplam
Erkek	18	38	56
Kadın	15	9	24
Toplam	33	47	80

dereceli $\chi 2$ dağılışında cetvel değerinden küçüktür ($P>0.05$). Bu iki-boyutlu tablolara ait üç-boyutlu tablonun (Tablo 2 a), 4 farklı log-linear modele göre beklenen değerleri Tablo 3'de verilmiştir. Tablo 3'de 3. sütun tamamen bağımsız modeli [1][2][3], 4. sütun hastalık durumu ile sakroiliitis durumunun ilişkili olduğu modeli [12] [3], 5. sütun hastalık durumu ve cinsiyet ile sakroiliitis durumu ve hastalık durumunun ilişkili olduğu modelini [13][23], 6. sütun tüm ikili ilişkilerin mevcut olduğu modelini [12][13][23] göstermektedir.

Tablo 3. Tablo 2 a'daki Verilerin 4 Log-Linear Model Altında Beklenen Değerleri.
Table 3. Estimated Expected Values Under Four Log-Linear Models For Entries in Table 2 a.

(1) Hücre (i,j,k)	(2) Gözlenen	(3) ([1][2][3])	(4) ([12][3])	(5) ([13][23])	(6) ([12][13][23])
(1, 1, 1)	17	12.1	16.8	9.6	14.3
(2, 1, 1)	1	11.0	6.3	8.4	3.7
(1, 2, 1)	13	17.3	12.6	20.4	15.7
(2, 2, 1)	25	15.6	20.3	17.6	22.3
(1, 1, 2)	7	5.2	7.2	7.5	9.7
(2, 1, 2)	8	4.7	2.7	7.5	5.3
(1, 2, 2)	5	7.4	5.4	4.5	2.3
(2, 2, 2)	4	6.7	8.7	4.5	6.7

Tablo 2 a'daki verilere uyan 8 log-linear model vardır. Bu modellere ait tahminlenmiş beklenen değerler (2.1), (2.11) ve (2.13) formüllerinden hesaplanabilir.

Tüm ikili ilişkilerin mevcut olduğu model için (Tablo 3'de 6. sütunda verilen model) direkt tahmin yapılamadığından bu modele ait beklenen değerler tekrarlı tahmin yöntemi ile elde edilmiştir. Bu yöntemin uygulanması Bircan (1990) tarafından gösterilmiştir.

Tablo 3'de verilen modeller ve diğer modellerle ilgili uyum istatistikleri χ^2 ve G^2; 2.18 ve 2.19 eşitliklerinden hesaplanarak Tablo 4'de gösterilmiştir.

Tablo 4'de görüldüğü gibi tüm log-linear modeller χ^2 dağılışına ait ilgili serbestlik derecesinde önemlidir. Tablo 2 a'daki verilerin tüm iki-boyutlu marjinal toplamlarına bağımsızlık şartı altında Pearson x^2 istatistiği uygulanması halinde; Tablo 2 c'ye ait değişkenlerin bağımsız Tablo 2 b ve d'ye ait değişkenlerin birbiri ile bağımlı olduğu bulunmuştu. Bu verilere log-linear modellerin uydurulması sonucunda ise, tüm ikili ilişkilerin ve ilave olarak üçlü ilişkinin de modelde yer alacağı gerekmektedir. Bu nedenle, eğer veriler üç-boyutlu çapraz-sınıflandırılabiliyorsa; o taktirde bu verilere iki-boyutlu marjinal toplamlarının analizi yerine log-linear modellerin uydurulması daha uygundur.

Tablo 4. Tablo 2 a'daki Verilere Ait Çeşitli Log-Linear Modeller İçin χ^2 ve G^2 Uyum İstatistikleri.
Table 4. Log Linear Models Fit to Data in Table 2 a, and Their Corresponding Goodness-of-Fit Statistics.

Model	Serbestlik Derecesi	χ^2	G^2
[12][13][23]	1	9.83*	9.89*
[12][13]	2	17.94*	17.86*
[13][23]	2	17.99*	20.98*
[12][23]	2	10.82*	11.59*
[12][3]	3	18.54*	17.94*
[13][2]	3	22.60*	27.33*
[23][1]	3	18.03*	21.06*
[1][2][3]	4	22.50*	27.42*

* 0.05 ihtimal seviyesinde, belirtilen serbestlik dereceli χ^2 dağılışında önemlidir.

Üç-boyutlu çapraz-sınıflandırılmış tabloların tüm iki-boyutlu marjinal toplamlarının klasik χ^2 testine tabi tutulması, üçüncü faktörün tesirinin elimine edilmesinden dolayı bilgi kaybına ve dolayısıyla da hatalı değerlendirmeye neden olmaktadır. Yine üç-boyutlu tabloların tüm iki-boyutlu marjinal toplamlarının analizleri eşanlı olarak yapılamamaktadır. Ayrıca bu şekildeki analizde, faktör sayısı üçden fazla ise; yüksek dereceden interaksiyonlar gözönüne alınamamaktadır. Üç-boyutlu çapraz sınıflanmış verilere loglinear modellerin uydurulması sonucunda ise, yukarıda verilen eksiklikler tamamıyla giderilmektedir.

Sonuç olarak; üç-faktörlü çapraz-sınıflandırılmış, eksik olmayan tabloların analizinde; loglinear modellerin uydurulması metodu ile, araştırıcının ihtiyaç duyduğu bilgileri temin edecek tarzda analiz yapılabilmektedir. Günümüzde bilgisayar imkanlarının fevkalade geniş olması, metodun uygulanmasındaki karmaşıklığıda ortadan kaldırmaktadır.

KAYNAKLAR

Barlett, M.S., 1935. Contingency tables interaction. J. Roy. Statist. Soc. Suppl. 2 : 248-249.

Bircan, H., 1990. İki ve Üç Boyutlu Çapraz-Sınıflandırılmış Verilere Log-Linear Modellerin Uydurulması. Yüksek Lisans Tezi. Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Erzurum.

Birch, M.W., 1963. Maximum likelihood in theree-way contingency tables. J. Roy. Statist. Soc. B25, 220-233.

Bishop, Y.M.M., 1967. Multidimensional contingency tables cell estimates. Ph. D. dissertation, Department of Statistics, Harvard University.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W.,1975. Discrete Multivariate Analysis. Theory and Practice, Mass, The MIT Press, 1975.

Cressie, N. and Read, T.R.C., 1989. Pearson's X^2 and the loglikelihood ratio statistics. G^2 A Comparative review. International Statist. Review 57, 19-43.

Darroch, J.N., 1962. Interaction in multi-factor contincency tables. J.Roy. Statist. Soc. Ser. B24, 251-163.

Fienberg, S.E., 1970a. An iterative procedure for estimation in contingency tables. Ann. Math. Statist. 41 : 907-917.

Fienberg, S.E., 1970b. The analysis of multidimensional contingency tables. Ecology 51, 419-433.

Fienberg, S.E., 1977. The analysis of Cross-Ciassified Categorical data Cambridge, Mass., The MIT Press.

Fisher, R.A., 1922. On the interpretation of chi-square from contingenay tables, and the calculation of P. J. Roy. Statist. Soc. 85 : 87-94.

Good, I.J., 1963. Maximum entropy for hypotheses formulation especially for multidimensional contingency tables. Ann. Math. Statist. 34, 911-934.

Goodman, L.A., 1963. On methods for comparing comtingency tables. J. Roy. Statist. Soc. A126, 94-108.

Goodman, L.A., 1964. Interactions in multidimensional contingency tables. Ann. Math. Statist. 35, 632-646.

Haberman, S.J., 1974. Loglinear models for frequency tables with ordered classifications. Biometrica. 30, 589-600.

Khamis, H.J. and Hinkelmenn, K., 1984. Log-linear-model analysis of the assocication between disease and genotype. Biometrics 40, 177-188.

Kılıç, V., 1984. 42 Behçet Olgusunda Klinik Bulgular ve Artritis'in Klinik, Radyolojik Özellikleri Uzmanlık Tezi. Atatürk Üniversitesi, Tıp Fakültesi, Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı, Erzurum.

Pearson, K., 1900. On a criterion that a given system of deviations from the probable in the cese of e carrelated system of variables in such that it can be reasonably supposed to hava arisen from random sampling. Philos. Mag. 50 (5) : 157-175.

Plackett, R.L., 1962. A note on interactions in contingency tables. J.Roy Statist Soc. B24, 162-166.

Roy, S.N. and Kastenbaum, M.A., 1956. On the hypothesis of no "interaction" in a multi-way contingency tables. Ann. Math. Statist. 27, 749-757.

STATGARAPHICS, 1988. Statistical Graphics System. By Statistical Graphics Corporation. User's Guide-Reference STSC, Inc.

Wilks, S.S., 1935. The likelihood test of independence in contingency tables. Ann. Math. Statist. 6 : 190-196.