

170246

ANKARA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TAMAMLANMAMIŞ İKİ YÖNLÜ  
OLUMSALLIK TABLOLARDA EN ÇOK OLABİLİRLİK TAHMİNİ

ANIL AKTAŞ

İSTATİSTİK ANABİLİM DALI

ANKARA  
2005

Her hakkı saklıdır

Yrd. Doç.Dr.Cemal ATAKAN danışmanlığında, Anıl AKTAŞ tarafından hazırlanan bu çalışma 07/10/2005 tarihinde aşağıdaki jüri tarafından İSTATİSTİK Anabilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Ayşen APAYDIN,  
Ankara Üniversitesi Fen Fakültesi



Üye :Yrd.Doç.Dr.Mehmet N.ORMAN,  
Ankara Üniversitesi Veterinerlik Fakültesi



Üye :Yrd.Doç.Dr.Cemal ATAKAN,  
Ankara Üniversitesi Fen Fakültesi



**Yukarıdaki sonucu onaylarım**

**Prof. Dr. Ülkü MEHMETOĞLU**  
**Enstitü Müdürü**

## ÖZET

Yüksek Lisans Tezi

### TAMAMLANMAMIŞ İKİ YÖNLÜ OLUMSALLIK TABLOLARDA EN ÇOK OLABİLİRLİK TAHMİNİ

Anıl AKTAŞ

Ankara Üniversitesi  
Fen Bilimleri Enstitüsü  
İstatistik Anabilim Dalı

Danışman: Yrd. Doç. Dr. Cemal ATAKAN

Eksik gözlemlerin bulunduğu olumsuzluk tabloları, tamamlanmamış tablo olarak adlandırılır. Bu eksik gözlemler iki türden dolayı oluşabilir. Bunlar örneklem sıfır ve yapısal sıfırdır. Örneklem sıfır, örneklem değişkenliğine bağlı iken, yapısal sıfır, bir hücrenin sıfır değerini almasıdır ve verinin yapısına bağlıdır. Tamamlanmamış iki yönlü olumsuzluk tablolarına ilişkin tahminler; olumsuzluk tablolarında bulunan eksik gözlemlerin tahminini amaçlar. Tamamlanmamış olumsuzluk tablolarındaki boş gözlemlerin tahmin yöntemlerinden biri, en çok olabilirlik yöntemidir. Tamamlanmamış iki yönlü tablolarında eksik gözlemlerin en çok olabilirlik tahmini için bazı yöntemler geliştirilmiştir. Bu çalışmada, bu yöntemlerden Poisson model yapısı göz önüne alınmıştır. Ayrıca eksik gözlemlere ilişkin elde edilen En çok olabilirlik tahminleri ve bu tahminler kullanılarak uyum iyiliği testi sonuçları verilmiştir.

**2005, 54 sayfa**

**ANAHTAR KELİMELEER:** Yapısal sıfır, Tamamlanmamış 2 yönlü olumsuzluk tablo, En çok olabilirlik tahmini, Tahmin yöntemleri

## ABSTRACT

Master Thesis

### MAXIMUM LIKELIHOOD ESTIMATES IN INCOMPLETE TWO-WAY CONTINGENCY TABLES

Anıl AKTAŞ

Ankara University  
Graduate School of Natural and Applied Sciences  
Department Statistics

Supervisor:Asst. Prof.Dr. Cemal ATAKAN

Contingency tables with missing data are called Incomplete tables. There are two different types of missing data . These are sampling zeros and structural zeros. Sampling zeros depend on sampling variability but structural zeros are known to have a zero value and depend on the structure of data. The analysis for incomplete tables depend on the estimation of empty cells. One of the methods used in estimation of empty cells is Maximum Likelihood Estimates. There are different methods that are improved to estimate Maximum likelihood Estimates. In this study, these methods are given with Poisson linear model. Also, Maximum Likelihood estimates and Goodness of fit test results are given for missing data.

**2005 , 54 pages**

**KEY WORDS:** Structural zeros, Two-way incomplete contingency tables, Maximum Likelihood Estimates and estimates methods

## TEŐEKKÜR

Bu alıőmada beni ynlendiren danıőman hocam, Sayın Yrd. Do.Dr. Cemal ATAKAN (Ankara niversitesi Fen Fakltesi İstatistik Blm)' a teőekkrlerimi sunarım.

Her zaman benim yanımda olan ve bana her bakımdan destek veren ve bana inanan canımdan ok sevdiėim aileme, sonsuz teőekkrler. Ayrıca, benden desteklerini esirgemeyen byklerime, arkadaőlarıma teőekkrler.

Anıl AKTAŐ

Ankara, Ekim 2005



## İÇİNDEKİLER

ÖZET ..	i
ABSTRACT ..	ii
TEŞEKKÜR.....	iii
SİMGELER DİZİNİ ..	v
ŞEKİLLER DİZİNİ ..	vi
ÇİZELGELER DİZİNİ ..	vii
1.GİRİŞ.....	1
2. OLUMSALLIK TABLO YAPILARI.....	5
2.1 İki Yönlü Olumsuzluk Tablolar.....	5
2.2 Örneklem Sıfır ve Yapısal Sıfır Kavramları.....	5
2.3.Tamamlanmamış İki Yönlü Olumsuzluk Tabloları.....	6
2.4.Quasi –Bağımsızlık.....	8
2.5.Tamamlanmamış Olumsuzluk Tabloların Birleştirilmesi Ve Ayrıştırılması .....	10
2.6 Tamamlanmamış Olumsuzluk Tablolar İçin Örneklem Yöntemleri .....	12
2.6.1.Poisson Örneklem Modeli.....	12
2.7. En Çok Olabilirlik Tahmin Yöntemi.....	16
3.TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDA EN ÇOK OLABİLİRLİK YÖNTEMLERİ VE UYUM İYİLİĞİ TESTİ .....	17
3.1.Newton-Raphson Yöntemi.....	17
3 2.EM Algoritması .....	20
3.2.1. EM algoritmasının formülasyonu.....	21
3.3. Quasi-Bağımsızlık Altında En Çok Olabilirlik Tahmini İçin İteratif Çözümler .....	26
3.3.1. Quasi-Bağımsızlık için doğrudan tahmin yöntemi.....	29
3.4.Kapalı –Form Tahminleri.....	35
3.4.1.Log-Doğrusal model yapısı ve tahminler .....	36
3.5 Tamamlanmamış İki Yönlü Olumsuzluk Tablolarda Uyum İyiliği Testi.....	40
4. UYGULAMA .....	41
5. TARTIŞMA ve SONUÇ.....	48

<b>KAYNAKLAR</b> .....	49
<b>EKLER</b> .....	50
<b>EK 1 Tamamlanmamış olumsuzluk Tablolarda Parametre Tahminlerinin SAS</b>	
<b>Komutları ve Çıktıları</b> .....	51
<b>EK 2 Tamamlanmamış Olumsuzluk Tablolarda Parametre Tahminlerinin STATA</b>	
<b>Çıktıları</b> .....	53
<b>ÖZGEÇMİŞ</b> .....	54



## SİMGELER DİZİNİ

EÇOB	En çok Olabilirlik Tahmini
EM	EM (Expectation-Maximization) algoritması





## ŞEKİLLER DİZİNİ

Şekil 3.1 Blok-üçgensel tabloların genel gösterimi.....	32
Şekil 3.2 Blok-basamak tabloların genel gösterimi.....	34



## ÇİZELGELER DİZİNİ

Çizelge 2.1	Eksik gözlemlerle 2 x 2 tablo yapısı .....	8
Çizelge 2.2	3 x 3 tamamlanmamış olumsuzluk tablosu yapısı .....	9
Çizelge 2.3	2 x 3 bağımsız olumsuzluk tablo .....	9
Çizelge 2.4	3 x 2 bağımsız olumsuzluk tablo. ....	10
Çizelge 2.5	3 x 3 tamamlanmamış olumsuzluk tablo .....	10
Çizelge 2.6	Ayrılabilir tablo yapısının gösterimi .....	11
Çizelge 2.7	İki ayrılmış tablo yapısının gösterimi .....	11
Çizelge 3.1	EM algoritmasının yapısını gösteren tablo bileşenleri.....	24
Çizelge 3.2	3x 3 biçiminde tamamlanmamış tablo .....	30
Çizelge 3.3	2x2 biçiminde tamamlanmış tablo .....	30
Çizelge 3.4	5x5 biçiminde tamamlanmamış tablo yapısı.....	31
Çizelge 3.5	Ayrılmış alt tablo yapısı .....	31
Çizelge 3.6	Blok-üçgensel tamamlanmamış alt tablo yapısı.....	32
Çizelge 3.7	Blok-üçgensel tamamlanmamış alt tablo yapısı.....	32
Çizelge 3.8	Eksik gözlemlerin bulunduğu 2x2 lik tablo yapısı.....	36
Çizelge 4.1	100 birimlik örnek için eksik gözlemleri içeren tamamlanmamış olumsallık tablo .....	41
Çizelge 4.2	100 birimlik örnek için eksik gözlemleri içermeyen tamamlanmamış olumsallık tablo .....	44
Çizelge 4.3	100 birimlik örnek için tamamlanmış olumsuzluk tabloya ilişkin uyum iyiliği test sonuçları .....	44
Çizelge 4.4	40 birimlik örnek için eksik gözlemleri içeren tamamlanmamış olumsallık tablo .....	45
Çizelge 4.5	40 birimlik örnek için eksik gözlemleri içermeyen tamamlanmamış olumsallık tablo .....	46
Çizelge 4.6	40 birimlik örnek için tamamlanmış olumsuzluk tabloya ilişkin uyum iyiliği test sonuçları .....	46

## 1. GİRİŞ

İstatistiksel veriler çok deęişik türde elde edilir. Farklı veriler için farklı analiz yöntemleri vardır. Bir çok analiz yöntemi sürekli veriler için geliştirilmiştir. Üzerinde gözlem aldığımız deęişkenler sürekli olmayabilir. Bazı olguların açıklanmasında kategorik deęişkenler kullanılır. Kategorik deęişken, sayılarla ifade edilmeyen ölçek deęişkenlerdir. Saç rengi, cinsiyet, öğrenim durumu, hastalık belirtisi gibi deęişkenler kategorik deęişkenlere ilişkin örneklerdir ve bu deęişkenler sayısal deęerler olmadığından dolayı analizlerde deęişkenlerin yapısına göre kodlama yapılır.

Kategorik deęişkenlere ilişkin verilerin analizlerinde olumsuzluk tabloları sıkça kullanılmaktadır. Farklı yapılarda olumsuzluk tabloları söz konusudur. Genel olarak, olumsuzluk tabloları tamamlanmış ve tamamlanmamış tablolar olarak ikiye ayrılabilir. Tamamlanmış tablolarda bütün birimlere ilişkin deęerler gözlenmesine rağmen, tamamlanmamış tablolarda bazı birimlere ilişkin deęerler bazı nedenlerden dolayı elde edilememektedir. Bu durumda eksik veri söz konusudur. Bu çalışmada tamamlanmamış olumsuzluk tabloları üzerinde durulacaktır.

Çalışmalarda amaca ilişkin veri toplarken bazı sebeplerden dolayı bazı birimlere ait gözlemler elde edilemeyebilir. Bu tür elde edilemeyen veri, kayıp veri olarak nitelendirilir.

Olumsuzluk tablolarla yapılan analizlerde, veri kaybı durumunda tablonun gözelerinde ki gözlem sayısı azalır. Veri kaybının birkaç nedeni olabilir. Veri kaybı; verinin yapısı gereęi (yapısal sıfır) ya da verinin kaydedilmesi esnasında (örneklem sıfır) olabilir. Böyle durumlarda kayıp veriye sahip gözeler eksik göze, o gözelere ait birimlere ilişkin deęerler sıfır olarak kabul edilir. Eksik verilerin bulunduğu olumsuzluk tablolar, tamamlanmamış tablolar olarak adlandırılır.

İki kategorik deęişkendenden oluşan tamamlanmamış tablolara iki yönlü olumsuzluk tabloları denir. İki yönlü olumsuzluk tablolarda, satır ve sütun deęişkenleri çok kategorili

(düzeyli) olabilir. Tamamlanmamış tablolarda yani satır ve sütun değişkenlerine ait bazı birimlerin değerlerinin gözlenmemesi durumunda elde edilen tablolarda hücre olasılıklarının tahminlerinin elde edilmesi her zaman kolay değildir.

Tamamlanmamış olumsuzluk tablolarda, boş gözlemlere ilişkin olasılıkların tahmin yöntemlerinden biri En çok olabilirlik(EÇOB) tahmin yöntemidir. En çok olabilirlik tahminleri, farklı örnekleme planlarından elde edilebilir. Bu örnekleme planlarından üçü; Poisson, basit çok terimli ve çarpımsal çoklu örnekleme planlarıdır. Bu üç örnekleme planı için, EÇOB tahminleri aynıdır. Örnekleme planları eksik veri varlığından etkilenmez. Bu çalışmada, En çok olabilirlik tahminleri Poisson modeline göre elde edilmektedir. Poisson model; iki bileşenden oluşmaktadır. Bu bileşenlerden rasgele bileşen, gözlenen sayıların ortak olasılık dağılımlarını tanımlarken, sistematik bileşen beklenen değer ve parametreler arasındaki bağıntı fonksiyonunu tanımlar. Tamamlanmamış tabloların EÇOB tahminlerinde Poisson model için tasarım matris yapısı kullanılarak olabilirlikler en büyük yapılır.

Tamamlanmamış tablolara ilişkin analizlerde eksik gözlemlerin tahmini için EÇOB tahminleri doğrudan kullanılmamasına rağmen, eksik gözlemlerin EÇOB tahminleri için bazı yöntemler geliştirilmiştir. Bunlar EM algoritması, Newton-Rapshon, kapalı-form tahmin yöntemleri ve quasi-bağımsızlık altında kullanılan EÇOB tahmin yöntemleridir.

En çok olabilirlik tahminlerin elde edilmesinde daha az yinelemeden dolayı EM algoritması daha yaygın kullanılmaktadır. Her yinelemede iki adım vardır. Bunlar; E ve M adımlarıdır. EM algoritmasında, eksik gözlemler için başlangıç değerler atandıktan sonra parametreler tahmin edilir. Bu parametre tahminleri kullanılarak parametreler için yeni tahmin değerleri elde edilir ve işlemler tek bir sonuca yaklaşıncaya kadar devam eder. Newton-Rapshon yöntemi de benzer bir yöntemdir fakat EM algoritmasına göre daha yavaş bir yöntem olduğu belirtilmektedir (McLachlan and Krishnan 1997). EM ve Newton-Rapshon yöntemlerinde için fazla sayıda yineleme kullanılırken, kapalı-form tahminlerinin elle hesaplanması mümkün olabilmektedir (Baker *et al.* 1992)

Literatürde tamamlanmamış olumsuzluk tabloların EÇOB tahminlerinin elde edilmesi üzerine çalışmalar bulunmaktadır. Chen and Fienberg (1974) çalışmalarında örneklem planlarının EÇOB tahminlerini ve uyum iyiliği testinin incelenmesini, doğum ağırlığı düşük olan bebeklerin serum ve sağlık indeks skorlarına göre kısmi sınıflandırılması durumundaki tahminlerini içeren bir uygulama ile vermişlerdir.

Fuchs (1982) çalışmasında tamamlanmamış tabloların EÇOB tahminlerini EM algoritması yardımıyla incelemiştir. Algoritmanın avantajları hakkında bilgi verirken, log-doğrusal modellerin uyum iyiliği testlerini yaşlı insanları koruyucu bir servisin etkinliğinin araştırılmasında EM algoritmasının kullanıldığı bir uygulama ile birlikte vermiştir.

Baker *et al.* (1992) çalışmalarında log-doğrusal modelleri ve kapalı-form tahminlerini kullanarak eksik gözlemlili verilere ilişkin olumsuzluk tablolarının analizini açıklamışlardır ve çalışmalarında sigara değişkeni ile bebeklerin doğum ağırlıklarının ilişkisini gösteren bir uygulama vermişlerdir.

Lipsitz *et al.* (1998) çalışmalarında tamamlanmamış verilere ilişkin istatistiksel yapıları, olabilirlikleri ve model yapılarını SAS ve STATA paket programları kullanarak vermişlerdir.

Çalışmada, tamamlanmamış olumsuzluk tablolarda boş gözlemlere ilişkin değerlerin tahmin edilmesi amaçlanmıştır. Bu tahmin değerlerinin de EÇOB tahmin yöntemleri kullanılarak incelenecektir. Tamamlanmamış iki yönlü olumsuzluk tabloları en çok olabilirlik tahminlerinin elde edilmesinin ve uyum iyiliği testinin incelendiği çalışmanın İkinci Bölümünde; tamamlanmamış tablo, tamamlanmamış iki yönlü olumsuzluk tablo yapıları, quasi-bağımsızlık, örneklem yöntemi, en çok olabilirlik tahmini hakkında genel bilgilere yer verilecektir.

Üçüncü Bölümde, simülasyonla üretilen verilerin yapısı ve bu verilerin SAS, STATA ve SPSS paket programlarında işleyişi ile tamamlanmamış olumsuzluk tablolarının en

çok olabilirlik tahmin yöntemleri ve uyum iyiliği testi hakkında bilgi verilecektir. Çalışma verilerinin istenilen analizleri ve sonuçların verildiği tablolar üzerinde durulacaktır.

Son bölümde de bulgularla ilgili olarak yorumlara ve sonuçlara yer verilecektir.



## 2. OLUMSALLIK TABLO YAPILARI

Yapılan çalışmalarda değişkenler kategorilerden oluştuğunda analiz yöntemlerinden biri, olumsuzluk tablolar ya da diğer ifade ile çapraz tablolar yardımı ile olmaktadır. Bazı durumlarda verilerin bir kısmı gözlenmemiş veya rapor edilememiş olabilir. Bu durumda eksik gözlemlerden oluşan tablolar ortaya çıkmaktadır. Bu tür tablolara, tamamlanmamış tablolar denir. Tamamlanmamış tablolara ilişkin analizlerin yapılabilmesi için eksik gözlemlerin bir şekilde tahmin edilmesi gerekir. Eksik gözlemlere ilişkin tahminlerden biri En çok olabilirlik tahmin yöntemidir. Çalışmada, iki yönlü olumsuzluk tablolar, örneklem sıfır ve yapısal sıfır kavramları, tamamlanmamış iki yönlü tablolar, quasi-bağımsızlık, tamamlanmamış olumsuzluk tabloların birleştirilmesi ve ayrıştırılması, bu tablolar için örneklem yöntemlerinden Poisson örneklem yöntemi ve tamamlanmamış tablolarda eksik gözlemlerin EÇOB tahminleri üzerinde durulacaktır.

### 2.1. İki Yönlü Olumsuzluk Tablolar

$R$  ve  $C$ , satır ve sütun sayıları olmak üzere iki kategorik değişkenden oluşan  $R \times C$  şeklindeki olumsuzluk tablolar iki yönlü olumsuzluk tablolar olarak adlandırılır.

Örneklem uzayı ;

$$X = \{(i,j) / 1 \leq i \leq R, 1 \leq j \leq C\}$$

olmak üzere;

$(i,j)$  hücresi için gözlenen sayı  $f(i,j)$  yada  $f_{ij}$  ile belirtilir.  $f_{i+}$  ve  $f_{+j}$ ;  $i$ -inci satır ve  $j$ -inci sütun için toplamları göstermektedir ve toplam örneklem genişliği;

$$N = \sum_{i=1}^R f_{i+} = \sum_{j=1}^C f_{+j}$$

dir. İki yönlü bir olumsuzluk tabloda  $R \times C$  tane hücre (göze) vardır (Rapallo 2003).

### 2.2. Örneklem Sıfır ve Yapısal Sıfır Kavramları

Tabloların birçok hücresi küçük sıklıklara sahipse, bu tablolar seyrek tablolar olarak adlandırılır.  $n$ ; hücre sayısı,  $N$ ; toplam örneklem genişliği olmak üzere  $(n/N)$  ifadesine

seyreklik indeksi denir. Örneklem genişliği küçük ya da örneklem genişliği büyük fakat hücre sayısı fazla olduğunda seyrek tablolar oluşur. Olumsuzluk tablolarda, iki tip boş hücreyle karşılaşılabilir. Bunlar örneklem sıfır ve yapısal sıfıra sahip hücrelerdir. Sıfır değerine sahip hücreler boş hücre olarak adlandırılır. Örneklem sıfır; örneklem çeşitliliğine ve hücre olasılıklarının küçüklüğü ile ilişkilidir. Örneklem genişliği arttıkça bu durum kaybolur (Agresti 1990).

Bir hücre her zaman sıfır gözlem değerine sahip ise yapısal sıfır olarak adlandırılır. Bu durum verinin yapısal bir sonucudur. Yapısal sıfırlı hücreler gözlemlerin kesilmesi veya rapor edilmemesi gibi birkaç farklı durumda oluşur. Bu durumda değişkenin seviyeleri birleştirilemez ve bu seviyelere ilişkin hücrelere sıfır olasılıkları atanır. Bu nedenle tamamlanmamış tablolarda, eksik gözlemlerin EÇOB tahminleri, sıfır değerine sahip olmayan gözlemlerin beklenen sıklıkları yardımıyla elde edilir (Bishop and Fienberg 1975).

### 2.3. Tamamlanmamış İki Yönlü Olumsuzluk Tablolar

Eksik verilerin bulunduğu olumsuzluk tablolara, tamamlanmamış tablo denir.  $X_{k1}$  ve  $X_{k2}$  ( $k = 1, 2, \dots, N$ ); iki kesikli rasgele değişken olmak üzere bu rasgele değişkenler  $N$  tane birim üzerinden gözlemlendiği kabul edilsin. Bu değişkenlerin ortak dağılımlarına ilişkin olasılıklar yani herhangi bir birimin  $(i, j)$ -inci hücreye ait olması olasılığı;

$$p_{ij} = P(X_{k1} = i, X_{k2} = j) \quad (i = 1, 2, \dots, R, j = 1, 2, \dots, C)$$

dir.

Aynı şekilde bu rasgele değişkenlere ilişkin marjinal olasılıklar sırasıyla;

$$p_{i+} = (X_{k1} = i), \quad p_{+j} = (X_{k1} = j)$$

dir. Bu olasılıklar bir birimin sırasıyla  $i$ -inci satır ve  $j$ -inci sütunda bulunması olasılıklarıdır.

$X_{k1}$  ve  $X_{k2}$  değişkenlerinin ortak olasılık fonksiyonu ;



$$f(x_{k1}, x_{k2} | p) = \prod_{i=1}^R \prod_{j=1}^C p_{ij}^{I[x_{k1}=i, X_{k2}=j]} \quad (2.1)$$

biçiminde ifade edilmektedir. Burada  $I[.]$  gösterge fonksiyonudur.

$R_{k1}, R_{k2}$  gösterge değişkenleri sırasıyla;

$$R_{k1} = \begin{cases} 1, & X_{k1} \text{ gözlendiğinde} \\ 0, & X_{k1} \text{ gözlenemediğinde} \end{cases}$$

ve

$$R_{k2} = \begin{cases} 1, & X_{k2} \text{ gözlendiğinde} \\ 0, & X_{k2} \text{ gözlenemediğinde} \end{cases}$$

biçiminde tanımlansın.

$X_{k1}$  ve  $X_{k2}$  değişkenlerini her ikisi de gözlendiğinde iki yönlü olumsuzluk tablosunda hücrelerdeki birim sayılarını belirten ifade

$$z_{ij} = \sum_{k=1}^N R_{k1} R_{k2} I[X_{k1} = i, X_{k2} = j] \quad (2.2)$$

biçimindedir.

$X_{k1}$  gözlendiğinde ve  $X_{k2}$  gözlenemediğinde;

$$w_{i+} = \sum_{k=1}^N R_{k1} (1 - R_{k2}) I[X_{k1} = i] \quad (2.3)$$

ve

$X_{k1}$  gözlenemediğinde ve  $X_{k2}$  gözlendiğinde ifade;

$$h_{+j} = \sum_{k=1}^N (1 - R_{k1}) R_{k2} I[X_{k2} = j] \quad (2.4)$$

dir (Lipsitz *et al.* 1998). (2.2), (2.3) ve (2.4) ifadeleri EÇOB tahminlerinin hesaplanmasında kolaylık sağlamaktadır. Rasgele değişkenin değeri gözlenemediğinde eksik gözlem söz konusudur. Eksik gözlemlere sahip olumsuzluk tablo yani tamamlanmamış bir olumsuzluk tablo Çizelge 2.1' deki biçimde ifade edilmiştir.

**Çizelge 2.1** Eksik gözlemlerle 2x2 tablo yapısı

$X_{k1}$	$X_{k2}$		Eksik
	1	2	
1	$z_{11}$	$z_{12}$	$w_{1+}$
2	$z_{21}$	$z_{22}$	$w_{2+}$
Eksik	$h_{+1}$	$h_{+2}$	$t_{++}$

#### 2.4. Quasi-Bağımsızlık

$N$  birimlik örneklem genişliğine sahip bir  $R \times C$  lik olumsuzluk tablonun  $(i, j)$ -inci hücresinin beklenen sıklığı  $m_{ij}$  olmak üzere; beklenen sıklıklardan oluşan bu tablonun satır, sütun ve toplam değerleri;

$$\begin{aligned} m_{i+} &= \sum_{j=1}^J m_{ij} \quad i = 1, 2, \dots, R \\ m_{+j} &= \sum_{i=1}^I m_{ij} \quad j = 1, 2, \dots, C \\ m_{++} &= \sum_{i=1}^I \sum_{j=1}^J m_{ij} \end{aligned} \quad (2.5)$$

biçiminde ifade edilir.

Satır ve sütunların tam bağımsız olduğu genel bir modelde;  $\{a_i\}$  ve  $\{b_j\}$  pozitif sabit değerler olmak üzere;

$$m_{ij} = a_i \cdot b_j \quad (i = 1, 2, \dots, R \quad j = 1, 2, \dots, C)$$

dir.

$R \times C$  olumsuzluk tabloda  $(i, j)$  hücresindeki bireylerin olasılığı  $p_{ij}$  ile gösterilsin. Bu durumda satır ve sütunların bağımsızlıkları,

$$p_{ij} = p_{i+} \cdot p_{+j} \quad (i = 1, 2, \dots, R \quad j = 1, 2, \dots, C) \quad (2.6)$$

biçiminde ifade edilir.

Tamamlanmamış iki yönlü tablolarda yapısal sıfır içermeyen tüm hücrelerin bir kümesi  $S$  ile gösterilsin ve tüm  $m_{ij} \in S$  olduğu varsayalım. Bu durumda  $S$  nin satır ve sütunları “quasi-bağımsız” olarak tanımlanır. Quasi-bağımsızlık; tablonun herhangi iki satırının (sütununun) ilişkili gözelerindeki birim sayılarının birbirlerine göre oranının aynı olmasıdır. Bu durumda bu satırlara (sütunlara) ilişkin gözelerdeki yapısal sıfırlar dikkate alınmaz. Bir anlamda quasi -bağımsızlık, bağımsızlığın özel bir biçimidir. Quasi bağımsızlık, karesel tablolar için geçerlidir. Köşegen hücreler haricindeki gözelerin bağımsızlığıdır (Rapallo 2001). 3x3'lük tablonun (1,1) gözesi unutulsun. Bu tablo Çizelge 2.2'de olduğu gibidir.

**Çizelge 2.2** 3 x 3 tamamlanmamış olumsuzluk tablo

-	$a_1b_2$	$a_1b_3$
$a_2b_1$	$a_2b_2$	$a_2b_3$
$a_3b_1$	$a_3b_2$	$a_3b_3$

Çizelge 2.2'de verilen tamamlanmamış tablonun birinci satırı atılarak 2x3 lük bağımsız olumsuzluk tablosu elde edilir. İlgili tablo Çizelge 2.3'de verilmiştir. Elde edilen bu 2x3 lük tablo tam bağımsız tablodur.

**Çizelge 2.3.** 2 x 3 bağımsız olumsuzluk tablo

$a_2b_1$	$a_2b_2$	$a_2b_3$
$a_3b_1$	$a_3b_2$	$a_3b_3$

Aynı şekilde Çizelge 2.2'de verilen tamamlanmamış 3x3'lük tablonun birinci sütunu atılarak Çizelge 2.4'te verilen 3x2'lik tablo elde edilir. Bu tabloda tam bağımsız tablodur.

**Çizelge 2.4.** 3 x 2 bağımsız olumsallık tablo

$a_1b_2$	$a_1b_3$
$a_2b_2$	$a_2b_3$
$a_3b_2$	$a_3b_3$

3 x 3'lük olumsallık tabloda üç köşegen de unutulduğunda, Çizelge 2.5'de verilen tablo oluşur. Bu tabloda satır veya sütunlardan birinin silinmesi boş hücreleri yok edemeyeceğinden, yeni bir alt tablo elde edilemez (Bishop and Fienberg 1975).

**Çizelge 2.5.** 3 x 3 tamamlanmamış olumsallık tablosu

-	$a_1b_2$	$a_1b_3$
$a_2b_1$	-	$a_2b_3$
$a_3b_1$	$a_3b_2$	-

## 2.5. Tamamlanmamış Olumsallık Tabloların Birleştirilmesi ve Ayrıştırılması

Tamamlanmamış olumsallık tabloların analizlerinde, kolaylık olması açısından eğer mümkün olabiliyorsa, tabloların birleştirilmesi veya ayrıştırılması durumu söz konusudur. Böyle oluşan tablolara ilişkin tahminlerin elde edilmesi de kolaydır.

Birleştirilmiş veya ayrıştırılmış iki yönlü olumsallık tablolarda, aynı satır veya sütundaki iki hücre yapısal sıfır içermiyorsa ilişkilidir. Yapısal sıfır olmayan hücrelerin bir kümesi veya bir hücre çifti yani ilişkili arka arkaya gelen iki eleman, hücrelerin zinciri ile bağlantılı olursa bağımlıdır. Bir hücre zincirinde ilişkili ardarda gelen elemanlar birbiriyle bağlantılı ise bağımlıdır. Sonuçta yapısal sıfır olmayan hücreler bir bağımlı kümeden geliyorsa, tamamlanmamış iki yönlü tablo bağımlıdır. Bir tamamlanmamış tabloda bağımlılık yoksa tabloya ayrılabilir tablo denir. Uygun satır ve sütun permütasyonundan sonra boş olmayan hücreler; bir hücrede olan bir birim değerinde olmayacak şekilde en azından iki ayrı alt tabloya bölünebilir. Bu tablolar, köşegenin altında bulunan ve en az iki tanesi üstüste gelmeyen bloklar içindeki boş

olmayan hücreler şeklinde düzenlenebilir. Bu tür tablolar köşegensel – blok tablo olarak adlandırılır.

Her blok bazı yapısal sıfırlar içerebilir. Yapısal sıfır ve beklenen sıklık ifadelerinde ayrılabilirlik durumu da tanımlanmaktadır. Gözlenmiş sayılar arasında rasgele sıfırlar olabilir. Gözlenen tablolarda sıfır olup olmamasını belirlemede amaç ayrılabilirliktir ve rasgele sıfır ile yapısal sıfır arasında bir fark yoktur. Eğer tamamlanmamış tabloda sıfır olmayan gözlemler birbirleriyle bağımlı değilse, gözlenen tabloların ayrılabilir olduğu söylenebilir (Bishop and Fienberg 1975). Ayrılabilir tablo yapısı Çizelge 2.6’ daki gibi ifade edilebilir.

**Çizelge 2.6.** Ayrılabilir tablo yapısının gösterimi

	$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	$B_6$
$A_1$	$X_1$	-	-	$X_2$	-	-
$A_2$	-	$X_3$	$X_4$	-	$X_5$	$X_6$
$A_3$	-	$X_7$	$X_8$	-	$X_9$	$X_{10}$
$A_4$	$X_{11}$	-	-	$X_{12}$	-	-

Çizelge 2.6’deki tablo yapısı yeniden düzenlenirse, Çizelge 2.7’deki gibi 2 ayrı alt tablo oluşur.

**Çizelge 2.7.** İki ayrılmış tablo yapısının gösterimi

	$B_1$	$B_4$	$B_2$	$B_3$	$B_5$	$B_6$
$A_1$	$X_1$	$X_2$	-	-	-	-
$A_4$	$X_{11}$	$X_{12}$	-	-	-	-
$A_2$	-	-	$X_3$	$X_4$	$X_5$	$X_6$
$A_3$	-	-	$X_7$	$X_8$	$X_9$	$X_{10}$

Tamamlanmamış olumsallık tablolardan, tamamlanmış olumsallık tablolar oluşturulabilir. Oluşan iki alt tablo tamamlanmış olumsallık tablosudur. Bu tip tablolara ilişkin gözlemlere ait birimlerin EÇOB tahminleri klasik EÇOB tahminleridir.

## 2.6. Tamamlanmamış Olumsallık Tablolar İçin Örneklem Yöntemleri

Tamamlanmamış olumsallık tablolarda EÇOB tahminleri, farklı örneklem planlarından elde edilmektedir. Beklenen göze sayıları ve olasılıklar için modellerin uygulanabilirliği ve örnekleme planlarının değişkenliği de model yapısı ile ilişki göstermektedir. Tamamlanmamış iki yönlü olumsallık tablo için modelin bağımsızlığı varsayımı altında gözlenen veriler, aşağıda verilen üç farklı örneklem planından hangisine sahip olursa olsun, beklenen göze sayıları için EÇOB tahmin edicisi aynıdır. Bu örnekleme planları;

1. Her hücre için bağımsız Poisson rasgele değişkeni için, Poisson örneklem planı,
2. Basit çok terimli örneklem planı,
3. Satırları ve sütunları bağımsız çok terimli bir kümede çarpımsal çoklu örneklem planıdır.

Bu üç plan aynı EÇOB tahminlerine sahiptir (Bishop *et al.* 1975). Bu çalışmada Poisson modeli üzerinde durulacağından diğer modellerin ayrıntıları verilmeyecektir.

### 2.6.1 Poisson Örneklem Modeli

Olumsallık tablolarda Poisson model yapısı, rasgele bileşen ve sistematik bileşen olmak üzere iki bileşenden oluşmaktadır. Rasgele bileşen, gözlenen sayıların ortak olasılık dağılımlarını tanımlarken; sistematik bileşen, beklenen değerler ve parametreler arasındaki bağıntı fonksiyonunu tanımlar. Beklenen değerler, parametrelerin bir fonksiyonudur.

Rasgele bileşen sayıların ortak dağılımlarını tanımlar.  $x_{ij}$ ,  $(i, j)$  gözesindeki gözlenen sayılar olmak üzere, beklenen değeri  $m_j$  olan Poisson dağılımına sahip olsun. Eğer  $\Pi^*$   $S^*$ 'de tamamlanmamış alt kümelerdeki tüm gözlemlerin çarpımını gösteriyorsa, bu durumda olabilirlik fonksiyonu;

$$L = \prod \frac{m_{ij}^{x_{ij}} e^{-m_{ij}}}{x_{ij}!} \quad i \neq j \quad (i = 1, 2, \dots, R \quad j = 1, 2, \dots, C) \quad (2.7)$$

dir.

Bu olabilirliğin doğal logaritması,

$$\log L = \sum \delta_{ij} [x_{ij} \log(m_{ij}) - m_{ij} - \log(x_{ij}!)] \quad (2.8)$$

şeklindedir. Burada,

$$\delta_{ij} = \begin{cases} 1 & ; \quad i = j \\ 0 & ; \quad i \neq j \end{cases} \quad (2.9)$$

dır.

Poisson dağılımı, olasılık dağılımlarının üstel ailesinin bir üyesidir. Gözlenen marjinal toplamlar, tüm  $i$  ler için  $x_{i+}$  ve tüm  $j$  ler için  $x_{+j}$  dir (Bishop *et al.* 1975).

Tamamlanmamış olumsuzluk tabloların EÇOB tahminlerine ilişkin analizlerde Poisson model için tasarım matris yapısı kullanılmaktadır. Poisson model yapısı kullanılarak olabilirlikler en büyük yapılabilir. Poisson model yapısının tasarım matris gösterimi aşağıdaki biçimde ifade edilmektedir.

$z_{ij}, w_{i+}, h_{+j}$  lar bağımsız Poisson rasgele değişkenleri olsun.  $z$ ,  $(RC \times 1)$ ;  $w$ ,  $(R \times 1)$  ve  $h$ ,  $(C \times 1)$  biçiminde vektörler olmak üzere;

$$\begin{aligned} z &= [z_{11}, \dots, z_{RC}]' \\ w &= [w_{11}, \dots, w_{R+}]' \\ h &= [h_{11}, \dots, h_{+C}]' \end{aligned} \quad (2.10)$$

dir.

$F$ , elemanları bağımsız Poisson rasgele değişkenlerden oluşan veri vektörü olmak üzere,

$$F = [z', w', h']$$

biçiminde ifade edilsin.

$p_{ij}$  'ler için EÇOB tahminleri  $F$  'in Poisson model yapısından elde edilebilir. Poisson model yapısı;

$$E(F) = Xp + \gamma \quad (2.11)$$

biçiminde verilmektedir.

Bu yapı  $X$  tasarım matrisi ve  $\gamma$  tamamlama vektörünü içermektedir.  $X$  ve  $\gamma$  nin boyutları tamamlanmamış tablonun boyutlarına bağlıdır. Modelde yer alan  $z$  değişkeni için model;

$$\begin{aligned} E(z_{ij}) &= z_{++} p_{ij} \quad (ij \neq RC); \\ E(z_{RC}) &= z_{++} (1 - \sum_{ij \neq RC} p_{ij}) = z_{++} - z_{++} \sum_{ij \neq RC} p_{ij} \end{aligned} \quad (2.12)$$

biçimindedir (Lipsitz *et al.* 1998)

$R=C=2$  için model yapısı;

$$E \begin{bmatrix} z_{11} \\ z_{12} \\ z_{21} \\ z_{22} \end{bmatrix} = z_{++} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ z_{++} \end{bmatrix}$$

ya da

$$E(z) = (z_{++} A_1) p + \gamma_1 \quad (2.13)$$

biçimindedir, burada  $A_1$  matrisi  $RC \times (RC-1)$  biçiminde olmak üzere ilk  $(RC-1)$  tane satırı bir birim matrisi vermekte ve son satırı ise  $-1$  lerden oluşmaktadır.  $\gamma_1$  vektörü  $(RC \times 1)$  biçiminde son satırı  $z_{++}$  ve diğer elemanları 0 olan tamamlama vektörüdür.

Sonuçta ;

$$E(z) = X_1 p + \gamma_1 \quad (2.14)$$

biçiminde ifade edilebilir. Burada  $X_1 = z_{++} A_1$  dir.

$X_{k1}$  satır ve  $X_{k2}$  sütun değişkenleri olmak üzere,  $X_{k1}$  gözlendiğinde ve  $X_{k2}$  gözlenemediğinde doğrusal model ifadeleri,

$$\begin{aligned} E(w_{i+}) &= w_{++} p_{i+} = w_{++} \sum_{j=1}^C p_{ij} \quad (i = 1, \dots, R-1); \\ E(w_{R+}) &= w_{++} (1 - \sum_{i=1}^{R-1} p_{i+}) = w_{++} \left[ 1 - \sum_{i=1}^{R-1} \sum_{j=1}^C p_{ij} \right] = w_{++} - w_{++} \left[ \sum_{i=1}^{R-1} \sum_{j=1}^C p_{ij} \right] \end{aligned} \quad (2.15)$$

biçiminde elde edilir.

$w_{++}$ , tamamlama vektörü olur.  $R=C=2$  için



$$\begin{aligned} \begin{bmatrix} E(w_{1+}) \\ E(w_{2+}) \end{bmatrix} &= w_{++} \begin{bmatrix} p_{1+} \\ p_{2+} \end{bmatrix} \\ &= w_{++} \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \end{bmatrix} + \begin{bmatrix} 0 \\ w_{++} \end{bmatrix} \end{aligned} \quad (2.16)$$

ya da

$$E(w) = (w_{++} A_2) p + \gamma_2 \quad (2.17)$$

olarak elde edilir. Burada  $A_2$  matrisi  $R \times (RC-1)$  biçiminde olmak üzere,  $R-1$  satırlı ve  $(R-1) \times C$  sütunlu blok-köşegen matrislere eşittir. Bloklar  $1 \times C$  lik 1'lerden oluşan bir vektördür.  $A_2$  nin son  $(C-1)$  sütunu tümüyle sıfırdır.  $A_2$  nin son satırı ilk  $(R-1)$  satırının toplamının negatifi ve ilk  $(R-1) \times C$  elemanları  $-1$  'e eşittir, son  $(C-1)$  elemanları 0 dir.  $\gamma_2$  vektörü  $(R \times 1)$  biçiminde kalan son satırı  $w_{++}$  dışındaki elemanları 0 olan tamamlama vektörüdür. Sonuçta ;

$$E(w) = X_2 p + \gamma_2 \quad (2.18)$$

biçiminde elde edilir. Burada  $X_2 = w_{++} A_2$  dir.

$X_{k1}$  gözlenemediğinde ve  $X_{k2}$  gözlendiğinde doğrusal model ifadesi;

$$\begin{aligned} E(h_{+j}) &= h_{++} p_{+j} = h_{++} \sum_{i=1}^R p_{ij} \quad (j = 1, \dots, C-1); \\ E(h_{+C}) &= h_{++} (1 - \sum_{j=1}^{C-1} p_{+j}) = h_{++} \left[ 1 - \sum_{i=1}^R \sum_{j=1}^{C-1} p_{ij} \right] = h_{++} - h_{++} \left[ \sum_{i=1}^R \sum_{j=1}^{C-1} p_{ij} \right] \end{aligned} \quad (2.19)$$

olarak elde edilir. Burada  $h_{++}$ , tamamlama vektörü olur.  $R=C=2$  için

$$\begin{aligned} \begin{bmatrix} E(h_{+1}) \\ E(h_{+2}) \end{bmatrix} &= h_{++} \begin{bmatrix} p_{+1} \\ p_{+2} \end{bmatrix} \\ &= h_{++} \begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \end{bmatrix} + \begin{bmatrix} 0 \\ h_{++} \end{bmatrix} \end{aligned} \quad (2.20)$$

ya da

$$E(h) = (h_{++} A_3) p + \gamma_3 \quad (2.21)$$

biçiminde verilir.

Burada  $A_3$  matrisi  $C \times (RC-1)$  biçiminde olmak üzere  $C-1$  satırlı  $R$  birim matrisin yatay kısmı gibi  $[I_{C-1} \ 0_{C-1}]$  şeklindedir.  $I_{C-1}$ ,  $C-1$  birim matris ve  $0_{C-1}$   $(C-1) \times 1$  lik 0 vektörüdür.

Son  $RC$ -inci sütun silinir.  $A_3$  ün son satırı ilk  $(C-1)$  satırının toplamının negatifidir.  $\gamma_3$  vektörü  $(Cx1)$  biçiminde son satırı  $h_{++}$  dışındaki elemanları 0 olan tamamlama vektörüdür. Sonuçta ;

$$E(h) = X_3 p + \gamma_3 \quad (2.22)$$

biçiminde elde edilir. Burada  $X_3 = h_{++} A_3$  dir.

Modelin genel yapısı; sırasıyla veri vektörü, tasarım matris ve tamamlama vektörü biçimindedir. Bu değerler sırasıyla;

$$F = \begin{bmatrix} z \\ w \\ h \end{bmatrix} \quad X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}$$

biçimindedir. Tasarım matrisi sabit içermez (Lipsitz *et al.* 1998).

## 2. 6. En Çok Olabilirlik Tahmin Yöntemi

Olumsuzluk tablolarının analizlerinde; gözlemlere ilişkin değerlerin tahminleri için farklı tahmin yöntemleri kullanılmaktadır. Bu tahminlerden biri EÇOB tahmin yöntemidir. EÇOB tahmini şu şekilde ifade edilebilir:

$X_1, \dots, X_n$ , birbirinden bağımsız rasgele vektörler ve ortak olasılık yoğunluk fonksiyonu  $g(x; \beta)$  olan kitleden bir örneklem olmak üzere;  $\beta$  nın olabilirlik fonksiyonu,  $L(\beta | X_1, X_2, \dots, X_n)$  biçiminde ifade edilmektedir. Olabilirlik fonksiyonu,

$$L(\beta | X_1, X_2, \dots, X_n) = g(X_1, X_2, \dots, X_n | \beta) = \prod_{i=1}^n f(X_i | \beta) \quad (2.23)$$

Olabilirlik fonksiyonunu maksimum yapan değer (örneklem bir fonksiyonu),  $\beta$  nın en çok olabilirlik tahmin edicisidir.  $\Theta$  parametre uzayı olmak üzere,  $\beta$  nın en çok olabilirlik tahmin edicisi (var olması durumunda)

$$\hat{\beta}_n(X_1, \dots, X_n) = \max_{\beta \in \Theta} L(\beta | X_1, \dots, X_n) \quad (2.24)$$

dır.

Genellikle, olabilirlik fonksiyonunun maksimize edilmesi yerine, fonksiyonun logaritması (log-olabilirlik) maksimize edilir. Bu fonksiyon,

$$l(\beta) = \log(L(\beta|X_1, \dots, X_n)) \quad (2.25)$$

dir (Roussas 1972).

Tamamlanmamış tablolara ait EÇOB tahmin problemlerini çözmek için bazı yöntemler geliştirilmiş. Bu yöntemlerden EM algoritması, Newton –Raphson algoritması, kapalı-form tahminleri, quasi-bağımsızlık altında Deming-Stephan ve bu yönteme eş değer olan yöntemlerle ilgili kısımlar bir sonraki bölümde bulunmaktadır.



### 3. TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDAN EN ÇOK OLABİLİRLİK TAHMİN YÖNTEMLERİ ve UYUM İYİLİĞİ TESTİ

Bu bölümde tamamlanmamış olumsuzluk tablolarına göre elde edilen modellerin parametre tahminlerine ilişkin yöntemler ve uyum iyiliği testi verilmektedir. Tamamlanmamış olumsuzluk tablolarında, klasik EÇOB yöntemlerin kullanılması kolay değildir. Eksik gözlemlerin bulunduğu olumsuzluk tablolarına ilişkin gözlemlerin EÇOB tahminlerinin elde edilebilmesi için bazı yöntemler geliştirilmiştir. Bu yöntemler Newton-Raphson, EM algoritması, quasi-bağımsızlık altında çeşitli tahmin yöntemleri ve kapalı-form tahminleridir.

#### 3.1 Newton-Raphson Yöntemi

Her zaman log olabirlik fonksiyonlarını en büyük yapmak mümkün olmayabilir. Bu tür durumlarda  $\beta$ 'nin EÇOB tahminlerini hesaplamadan yollarından biri de Newton-Raphson algoritması ve türevleri ile yapılabilir. Bu yöntem varyans tahminlerinin yanında hücre olasılıklarının EÇOB tahminleri ile ilgili bir yöntemdir.

$$\{S(x; \beta) = \partial \log L(\beta) / \partial \beta\} = 0 \quad (3.1)$$

olabirlik eşitliklerinin çözümü ile Newton-Raphson yöntemi  $\beta$  için uygun  $\beta^{(r)}$  belirlemede doğrusal Taylor serileri ile log-olabirlik fonksiyonu  $\log L(\beta)$  nin eğim vektörü  $S(x; \beta)$  ya yakınlaştırır.

$$S(x; \beta) \approx S(x; \beta^{(r)}) - I(\beta^{(r)}; x)(\beta - \beta^{(r)}) \quad (3.2)$$

dir.

Bu eşitliğin sağ tarafının sıfır olması ile yeni bir  $\beta^{(r+1)}$  adımı

$$\beta^{(r+1)} = \beta^{(r)} + I^{-1}(\beta^{(r)}; x)S(x; \beta^{(r)}) \quad (3.3)$$

elde edilir (McLachlan and Krishnan 1997).

Eğer log olabirlik fonksiyonu konkav ve tek modelli ise tek adımda  $\beta$ 'nin EÇOB tahminine yaklaşır ve ayrıca log-olabirlik fonksiyonu da  $\beta$ 'nin karesel fonksiyonu ise  $\{\beta^{(r)}\}$  nin yineleme zinciri ile  $\beta$  nin EÇOB tahminine yaklaşır. Log-olabirlik

fonksiyonu konkav değilse Newton-Rapshon yönteminin keyfi başlangıç değeri ile bir noktaya ulaşması garanti değildir.  $L(\beta)$  deki bu sebepler ve yeterli doğrulukta başlangıç değeri varsayımları altında  $S(x; \beta)$  nin  $\beta^*$  çözümü için Newton-Rapshon yöntemi genel karesel fonksiyon ile  $\beta^{(r)}$  yineleme zinciri ile ulaşılır.  $\Omega$  da  $\|\cdot\|$  normu ile uygundur. Öyle ki eğer  $\beta^{(0)}$  yeterli miktarda  $\beta^*$  ya yaklaşırsa

$$\|\beta^{(r+1)} - \beta^*\| \leq h \|\beta^{(r)} - \beta^*\|^2 \quad (3.4)$$

elde edilir.

(3.4) ifadesinde eşitliği sağlayacak sabit bir  $h$  değeri vardır. ( $r=0,1,2 \dots$ ) Karesel fonksiyon çok hızlıdır ve bu yöntem çok geniş boyuttaki işlemleri de göz önünde bulundurur. Fakat bu yöntemde bazı sorunlar bulunmaktadır. İlk olarak, her adımda  $d \times d$  biçiminde  $I(\beta^{(r)}; x)$  bilgi matrisinin hesaplanmasını (yani Hessian matrisinin negatifi) ve  $d$  doğrusal eşitlik sisteminin çözümünü gerektirir. Bu durumda genelde  $O(d^3)$  gibi aritmetik operasyonlarla çözüme kavuşturulabilir. Diğer bir sorun da (3.1) eşitliğinin çözümünde  $\{\beta^{(r)}\}$  yineleme zinciri için başlangıç değerinin belirlenmesinde pratik olmayan durumlar olmasıdır. Eyer noktası ve lokal maksimum yanı sıra lokal minimuma doğru eğilim gösterir. Bazı durumlarda, Böhninig ve Lindsay'a göre Newton-Rapshon yönteminin monoton olarak değiştirilebildiğini göstermişlerdir (McLachlan *et al.* 1997).

Newton-Rapshon yöntemi her  $r$  yinelemede  $I(\beta^{(r)}; x)$  nin değerlendirilmesini gerektirir. Bu durumda  $I^{-1}(\beta^*; x)$  gözlenmiş bilgi matrisinin tersi aracılığı ile  $\beta^*$  nin sınırlı değerlerinin kovaryans matrisinin tahminini sağlar. Eğer başlangıç değeri  $\beta$ 'nin  $\sqrt{n}$  - uygun tahmin edicisi ise, ilk adım yinelemeyi  $\beta^{(1)}$ , asimptotik olarak  $\beta$ 'nin yeterli tahmin edicisidir (McLachlan *et al.* 1997).

Tamamlanmamış tablolarda beklenen hücre frekansları için EÇOB elde edilmesinin diğer bir yöntemi EM algoritmasıdır. Fuchs (1982), çapraz tabloların bir veya birden

fazla deęişken deęerleri eksik olduęunda, log-doęrusal modeldeki parametrelerin EÇOB tahminlerini bulmak için EM algoritmasını kullanmıştır.

### 3.2. EM Algoritması

Birçok çalışmada bir veya birden fazla deęişkenin deęerleri orijinal örneklemin altkümeleri için eksik olabilir. Tamamlanmamış verilere ait log-doęrusal modellerin parametreleri için EÇOB tahminleri elde etmek genelde bir problem olmuştur. EM algoritması bu tür sorunları çözmek için uygun yöntemlerden biri olarak önerilmektedir.

EM algoritması, eksik veri problemlerde çok kullanılan bir yöntemdir. Ayrıca EÇOB tahminlerin hesaplanmasında iteratif hesaplamalar kullanan bir yöntemdir. EM algoritması, eksik gözlemlerin tahmini için ad-hoc yaklaşımı ile ilişkilidir öyle ki eksik gözlemler için birincil deęerler atandıktan sonra parametreler tahmin edilir. Bu birincil parametre tahminleri kullanılarak onların tahmini deęerleri ile adımlar tekrarlanır. Bu işlemler tek bir sonuca yaklaşıncaya kadar devam eder (McLachlan *et al.* 1997).

EM algoritması, yorucu olmayan ve hızlı hesaplama yapan bir yöntem olduğundan istatistiksel çalışmalarda standart bir araç olmuştur. EM algoritması, EÇOB tahminlerini ve tamamlanmamış verilerle ilgili problemlerde çok genel bir yineleme yöntemi olarak elde edilmiş. Algoritma E-adım ve M- adım olmak üzere iki adımdan oluşmaktadır (Dempster *et al.* 1974).

E- adımı, log olabilirlikleri verilen geçerli parametre tahminlerinin beklenenlerini bulmayı kapsar. M-adımı da beklenen log olabilirlikleri en büyük olanlar ile geçerli parametre tahminlerinin güncellenmesini kapsar.

#### 3.2.1 EM Algoritmasının Formülasyonu

Bir  $y$  gözlem vektörü için bu  $y$  gözlem serisi  $Y$  rasgele vektörünün gözlenen deęeri olsun.  $Y$  nin olasılık yoğunluk fonksiyonu;

$$g(y; \beta), \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}$$

biçiminde verilsin.

Bu fonksiyon parametre uzayındaki ( $\Omega$ ) bilinmeyen parametrelerin bir vektörüdür.  $\beta$  için  $y$  verisine ilişkin olabirlik fonksiyonu;

$$L(\beta) = g(y; \beta), \beta \in \Omega$$

dir.

Bir çok veri kümesi için En çok olabirlik tahminlerini kolayca elde etmek zor veya mümkün değildir. Gözlenen  $y$  vektörü, gözlenemeyen bir  $z$  vektörü ile bir  $x$  vektörünü oluştursun. Gözlenen  $y$  vektörü,  $x$  veri vektörünün gözlenebilir bir fonksiyonunu oluşturmaktadır[  $x = (y, z)$ ]. Bu  $x$  vektörü tam veri veya tam veri vektörü olarak adlandırılır. Gözlenmiş  $y$  tam olmayan veri vektörünün bir  $x$  tam veri vektörü ile birlikte düşünüp ele almanın yararı,  $\beta$  için  $x$  veri vektörüne dayalı,

$$L_c(\beta) = g_c(x; \beta), \beta \in \Omega$$

olabirlik fonksiyonundan;  $\hat{\beta}$  en çok olabirlik tahmin edicisinin elde edilmesidir.  $x$  tam veri vektörünün log-olabirlik fonksiyonu ;

$$\ln L_c(\beta) = \ln g_c(x; \beta) \quad (3.5)$$

dir (McLachlan *et al.* 1997).

$x$  veri vektörü içinde gözlenemeyen kısımlar olduğundan, en çok olabirlik tahminleri kullanmak tam olarak doğru değildir. Bu nedenle  $\ln L_c(\beta)$  fonksiyonu yerine bir  $\beta^{(0)}$  değeri için  $E_{\beta^{(0)}}[\ln L_c(\beta) / y]$  koşullu beklenen değeri alınmaktadır.  $E_{\beta^{(0)}}[\ln L_c(\beta) / y]$  koşullu beklenen değeri  $\beta$ 'nin bir fonksiyonu olmak üzere, bu koşullu beklenen değeri en büyük yapan,

$$\beta^{(1)} = \arg \max E_{\beta^{(0)}}[\ln L_c(\beta) / y] \quad (3.6)$$

değeridir.

Bu değer ile  $E_{\beta^{(1)}}[\ln L_c(\beta) / y]$  koşullu beklenen değeri yeniden bulunmakta ve

$$\beta^{(2)} = \arg \max_{\beta^{(1)}} [E_{\beta^{(1)}} [\ln L_c(\beta) / y]] \quad (3.7)$$

hesaplanmaktadır. Dempster *et al.* (1997) tarafından bu şekilde oluşan  $\beta^{(0)}, \beta^{(1)}, \beta^{(2)}, \dots$  dizisinin en çok olabilirlik tahmin edicisine yakınsadığını ve  $L(\beta^{(0)}), L(\beta^{(1)}), L(\beta^{(2)}), \dots$  dizisinin azalmayan bir dizi olduğunu ispatlamışlardır (McLachlan *et al.* 1997).

EM algoritmasında biçimsel olarak iki örneklem uzayı bulunmaktadır. Bunlar, gözlenen ve tam olmayan veri olarak isimlendirilen gözlemlerin bulunduğu  $\mathcal{Y}$  kümesi ile tam-veri olarak isimlendirilen verilerin bulunduğu  $\mathcal{X}$  kümesidir.  $\mathcal{X}$  örneklem uzayındaki  $x$  tam-veri vektörü yerine,  $\mathcal{Y}$  örneklem uzayındaki  $y = y(x)$  tam olmayan- veri vektörü gözlenmektedir.  $\beta$  için  $y$  verisine dayalı olabilirlik fonksiyonu,

$$L(\beta) = g(y; \beta), \beta \in \Omega \quad (3.8)$$

ve

$\beta$  için  $x$  vektörüne dayalı olabilirlik fonksiyonu,

$$L_c(\beta) = g_c(x; \beta), \beta \in \Omega$$

dır.

$$\mathcal{N}_y = \{x : y = y(x)\} \subset \mathcal{X}$$

olmak üzere,

$$g(y; \beta) = \int_{\mathcal{N}_y} g_c(x; \beta) dx \quad (3.9)$$

dır.

$L(\beta) = g(y; \beta)$  olabilirlik fonksiyonu  $y$  tam olmayan-veri vektörüne karşılık gelen  $Y$  rasgele vektörünün olasılık (yoğunluk) fonksiyonu ve  $L_c(\beta) = g_c(x; \beta)$  olabilirlik fonksiyonu  $x$  tam-veri vektörüne karşılık gelen  $X$  rasgele vektörünün olasılık (yoğunluk) fonksiyonudur.  $L(\beta)$  tam olmayan veri olabilirlik fonksiyonu,  $L_c(\beta)$  tam-veri olabilirlik fonksiyonudur.

Uygun bir  $\beta^{(0)}$  seçilir ve  $r=0$  alınır.



**E-ADIMI:** Bu adımda

$$Q(\beta; \beta^{(r)}) = E_{\beta^{(r)}} [\ln L_c(\beta) / y] \quad (3.10)$$

fonksiyonu elde edilir.

**M-ADIMI:** Bu adımda da

$Q(\beta; \beta^{(r)})$  fonksiyonu  $\Omega$  üzerinden en büyük yapılır, her  $\beta \in \Omega$  için,

$$Q(\beta^{(r+1)}, \beta^{(r)}) \geq Q(\beta; \beta^{(r)}) \quad (3.11)$$

olacak şekilde  $\beta^{(r+1)}$  bulunur.

Seçilen  $\varepsilon > 0$  için,

$$L(\beta^{(r+1)}) - L(\beta^{(r)}) > \varepsilon \quad (3.12)$$

ise  $r$  yerine  $r+1$  alınır ve E-adımına gidilir, aksi halde  $(\beta^{(r+1)})$  alınır.

EM algoritmasının her adım sonunda  $L(\beta)$  değeri azalmayandır.

$$L(\beta^{(r+1)}) > L(\beta^{(r)}) \quad r = 0, 1, 2, \dots$$

olmaktadır. Buna göre,  $L(\beta^{(0)}), L(\beta^{(1)}), L(\beta^{(2)}), \dots$  dizisi üstten sınırlı olduğundan yakınsama,

$$\beta^{(r)} \xrightarrow{r \rightarrow \infty} \hat{\beta}$$

sağlanmaktadır.

Üstel aileden dağılımlara uygulandığında, EM algoritması değişiklik göstermektedir.

E-adımı, tamamlanmış verilerin yeterli istatistikleri verildiğinde. gözlenmiş veri ve geçerli parametrelerin tahminini,

M-adım ise verilen yeterli istatistikler için olabilirlik eşitlikler için çözüm olarak yeni tahminler verir.

Yani ;

$$L(\beta) = g(x; \beta) = b(\beta) e^{\beta^T(x)h(x)} \quad (3.13)$$

dir.

$t(X)$  yeterli istatistiği için algoritma yapısı şu şekildedir:

i)  $r=0$  alınarak;  $\beta^{(0)}$  seçilir.

ii) E adımında;  $E_{\beta^{(r)}} [t(X) / y]$  elde edilir

iii) M adımında da  $E_{\beta[t(X)]} = E_{\beta^{(r)}}[t(X) / y]$  denklemi çözümü olan  $\beta^{(r+1)}$  bulunur.

iv)  $L(\beta^{(r+1)}) - L(\beta^{(r)}) > \varepsilon$  ( $\varepsilon > 0$ ) ise  $r$  yerine  $r+1$  alınır, E adımına gidilir, aksi halde  $(r+1)$  deki değer atılır (McLachlan *et al.* 1997).

$Y_1$  ve  $Y_2$  örneklem genişlikleri  $N$  olan iki kategorili değişkenler olsun. Kategoriler sırasıyla 1 ve 2 değerini alsın. Birinci grubun her iki birimi  $Y_1$  ve  $Y_2$  gözlenmiş olsun. İkinci grubun birimleri için  $Y_1$  gözlenmiş ve  $Y_2$  eksik olsun. Üçüncü grubun birimleri için de  $Y_1$  eksik ve  $Y_2$  gözlenmiş olsun. Bu durumda veriler 3 alt gruba ayrılmış gibi düşünülür (Ming-Yi 1999). Çizelge 3.1'de EM algoritmasının yapısını gösteren tablo bileşenleri verilmiştir.

**Çizelge 3.1** EM algoritmasının yapısını gösteren tablo bileşenleri

		$Y_2$		$Y_1$ için		$Y_2$ için	
		1	2				
$Y_1$	1	$m_{11}$	$m_{12}$	$\gamma_{21}$			
	2	$m_{21}$	$m_{22}$	$\gamma_{22}$	$\gamma_{31}$	$\gamma_{32}$	

$m = \{m_{ij}\}$  tamamlanmış durumlar için gözlenmiş hücre frekanslarını,  $\{\gamma_{wv}\}$   $w$ -inci eksik gözleme karşın  $v$ -inci marjinal hücre değerlerini ve  $r$ ; yineleme sayısını gösterebilir. ( $v=1,2, \dots, V_w$ ;  $w=1, 2, \dots, W$ )  $V_w$ ,  $w$  deki eksik gözlemlere ait hücrelerin kümesini gösterebilir.  $p$ ; hücre olasılıklarını belirtebilir.

E-adımı için;

$$\begin{aligned}
 \hat{n}_{11}^r &= m_{11} + \gamma_{21} \frac{\hat{p}_{11}^r}{\hat{p}_{11}^r + \hat{p}_{12}^r} + \gamma_{31} \frac{\hat{p}_{11}^r}{\hat{p}_{11}^r + \hat{p}_{21}^r}, \\
 \hat{n}_{12}^r &= m_{21} + \gamma_{21} \frac{\hat{p}_{12}^r}{\hat{p}_{11}^r + \hat{p}_{12}^r} + \gamma_{32} \frac{\hat{p}_{12}^r}{\hat{p}_{12}^r + \hat{p}_{22}^r}, \\
 \hat{n}_{21}^r &= m_{21} + \gamma_{22} \frac{\hat{p}_{21}^r}{\hat{p}_{21}^r + \hat{p}_{22}^r} + \gamma_{31} \frac{\hat{p}_{21}^r}{\hat{p}_{11}^r + \hat{p}_{21}^r}, \\
 \hat{n}_{22}^r &= m_{22} + \gamma_{22} \frac{\hat{p}_{22}^r}{\hat{p}_{21}^r + \hat{p}_{22}^r} + \gamma_{32} \frac{\hat{p}_{22}^r}{\hat{p}_{12}^r + \hat{p}_{22}^r},
 \end{aligned} \tag{3.14}$$

ifadeleri ve

M-adımı için;

$$\hat{p}_j^{g+1} = \hat{n}_j^g / N \quad i=1,2 \quad j=1,2 \tag{3.15}$$

$$N = \sum_{i,j=1}^2 m_{ij} + \sum_{k=1}^2 \gamma_{2k} + \sum_{k=1}^2 \gamma_{3k}$$

ifadeleri söz konusudur.  $\{\hat{p}_{ij}^0, i, j = 1,2\}$  hücre olasılıklarının başlangıç değeri olarak

$$\left\{ \frac{m_{ij}}{\sum_{i,j} m_{ij}}; i, j = 1,2 \right\}$$

hücre oranı alınır (Ming-Yi 1999).

EM algoritması, tamamlanmış verinin bir örnekleme için EÇOB tahminlerinin hesaplanması ile yeterli istatistiklerin değerlerinin düzeltilmiş yorumunu bulundurmaktadır. M-adımda tamamlanmış veri ile örneklemlerden elde edilen EÇOB tahminler için uygun algoritmalar kullanılabilir (Fuchs 1982).

EM algoritmasının diğer iteratif algoritmalar Newton-Rapshon ve Fisher skor yöntemlerine göre daha cazip özellikleri bulunmaktadır. Bunlardan bazıları:

1. EM algoritması her EM yinelemenin artan olabilirliği ile sayısal olarak sabittir.
2. Uygun koşullar altında, EM algoritması, güvenilir bir yaklaşıma sahiptir. Yani, parametre uzayından keyfi bir başlangıç noktası seçildiğinde yaklaşım, başlangıç

noktasının kötü seçimi veya log-olabilirlik fonksiyonundaki bazı bozuklar dışında maksimum olur.

3. EM algoritması kolayca hesaplanır, çünkü E-adımının her yinelemede yalnızca tam-verinin koşullu dağılımı altında beklenen değeri ve M-adımının her yinelemede ise tam-verinin EÇOB tahminlerini gerektirir.
4. EM algoritması genel olarak kolay bir programdır öyle ki ne olabilirliğinin değerlendirilmesi ne de onun türevlerini gerektirir.
5. Diğer yöntemlere göre daha kolay analitik çalışma gerektirir şöyle ki yalnızca tam-veri için log-olabilirliklerin koşullu beklenen değerlerini maksimum yapmak gerekmektedir.
6. Olabilirliklerdeki monoton artışı izlemek daha kolaydır.

Bunların yanı sıra algoritmanın eksik yanları da bulunmaktadır. Newton yöntemleri gibi global maksimuma yaklaşımı garanti değildir. Parametre tahminlerinin kovaryans matrisinin tahmininin oluşturulmasında sorun vardır. Bazı problemler için, E-adımı analitik olarak anlaşılabilir (McLachlan 1998).

### 3.3 Quasi - Bağımsızlık Altında En Çok Olabilirlik Tahmin İçin İteratif Çözümler

Quasi –bağımsızlık modelinin çarpımsal formunda, sıfır olmayan gözlemlere ilişkin beklenen sıklıklar için EÇOB tahminleri hesaplanırken, quasi bağımsızlık altında iteratif çözümler ile ilişkili olarak iki yöntem bulunmaktadır.

Quasi bağımsızlık modelinde;

$$m_{ij} = \delta_{ij} a_i b_j \quad i = 1, 2, \dots, R; j = 1, 2, \dots, C$$

dir.

$S$ ; yapısal sıfır içermeyen hücrelerin kümesi olmak üzere;

$$\delta_{ij} = \begin{cases} 1 & ; (i, j) \in S \\ 0 & ; d.h \end{cases} \quad (3.16)$$

dir.

Eğer  $m_{ij}$  için tek EÇOB tahmini varsa, bu tahminler ; quasi bağımsızlık ile ilişkili olarak  $R+C$  tane en büyük olabilirlik eşitlikleri tarafından belirlenmektedir. Bu eşitlikler;

$$\begin{aligned} \hat{m}_{i+} &= x_{i+} \\ \hat{m}_{+j} &= x_{+j} \end{aligned} \quad (i=1,2,\dots,R; j=1,2,\dots,C) \quad (3.17)$$

dir.

İlk yöntem göze değerlerinin beklenen sıklıklarının tahminlerini hesaplamak için Deming-Stephan adımsal oran uyumu ile tanımlanır. Bu yöntem için algoritma yapısı şu şekildedir:

0.adımda;

$$\hat{m}_{ij}^{(0)} = \delta_{ij}$$

$v$ -inci adımda;

$$\begin{aligned} \hat{m}_{ij}^{(2v-1)} &= \frac{\hat{m}_{ij}^{(2v-2)} x_{i+}}{\sum_j \hat{m}_{ij}^{(2v-2)}} \\ \hat{m}_{ij}^{(2v)} &= \frac{\hat{m}_{ij}^{(2v-1)} x_{+j}}{\sum_i \hat{m}_{ij}^{(2v-1)}} \quad \forall i, j \quad v = 1, 2, \dots \end{aligned} \quad (3.18)$$

dır.

Bu adım istenilen doğruluğa erişilene kadar devam eder (Bishop and Fienberg 1975). İkinci yöntem quasi-bağımsızlık altında çarpımsal form biçimindeki modellerin beklenen sıklıkların tahminlerinin hesaplanmasıdır Bu hesaplamada; iki adımsal yol vardır.

İlk adımsal yol;  $\{m_{ij}\}$  için EÇOB tahminlerin oluşturulmasında  $\{a_i\}$  ve  $\{b_j\}$  için en çok olabilirlik tahminler üretmektir. Quasi bağımsızlık model altında;

$$m_{ij} = \delta_{ij} a_i b_j \quad i = 1, 2, \dots, R, \quad j = 1, 2, \dots, C$$

şeklinde ifade edilmiştir.

Model  $R+C$  tane bağımsız parametreye sahiptir fakat sadece  $R+C-1$  tanesi için bağımsızlık görülebilir.

Bu durumda;

$$\sum_{i=1}^R \hat{m}_{i+} = \sum_{j=1}^C \hat{m}_{+j} = x_{++} \quad (3.19)$$

dır.

EÇOB tahmin için  $R+C-1$  tane denklem bulunmaktadır.  $\{a_i\}$  ve  $\{b_j\}$  nin çarpımsal sabitinin tek olduğu görülebilirse;  $R+C-1$  tane bağımsız parametre olduğu görülebilir.

Quasi bağımsızlığı kullanarak EÇOB tahmin eşitliği yazılırsa;

$$\begin{aligned} \hat{a}_i \sum_{j=1}^J \delta_{ij} \hat{b}_j &= x_{i+} \quad i = 1, 2, \dots, R \\ \hat{b}_j \sum_{i=1}^I \delta_{ij} \hat{a}_i &= x_{+j} \quad j = 1, 2, \dots, C \end{aligned} \quad (3.20)$$

$$\hat{a}_i = \frac{x_{i+}}{\sum_j \delta_{ij} \hat{b}_j} \quad \text{ve} \quad \hat{b}_j = \frac{x_{+j}}{\sum_i \delta_{ij} \hat{a}_i} \quad i = 1, 2, \dots, R; j = 1, 2, \dots, C$$

şeklinde elde edilir.

$\{a_i\}$  ve  $\{b_j\}$  tahminleri için ikinci adımsal yol:

0.adım;

$$b_j^{(0)} = 1 \quad j = 1, \dots, C \quad (v \geq 1)$$

$v$ -inci adım;

$$a_i^{(v)} = \frac{x_{i+}}{\sum_j \delta_{ij} b_j^{(v-1)}} \quad i = 1, \dots, R \quad (3.21)$$

$$b_j^{(v)} = \frac{x_{+j}}{\sum_i \delta_{ij} a_i^{(v)}} \quad j = 1, \dots, C$$

son adım;

$$\hat{m}_{ij}^{(2v)} = \delta_{ij} a_i^{(v)} b_j^{(v)} \quad \forall (i, j)$$

dir.

Yineleme değerleri bir değere ulaşıncaya kadar devam eder.  $m_{ij}$  nin tahmini EÇOB tahminidir. Bu iki yöntem eşdeğerdir. Deming-Stephan yöntemi ile bu değerlerin tahminlerine ulaşılmasına rağmen ikinci yöntem  $\{a_i\}$  ve  $\{b_j\}$  parametrelerinin tahmini

için daha uygundur. Bu yöntem ilkinde göre daha az hesaplama gerektirmektedir. Bu yöntemle her döngüden sonra tahminlerdeki değişim kontrol edilebilir (Bishop and Fienberg 1975).

### 3.3.1 Quasi Bağımsızlık İçin Doğrudan Tahmin Yöntemi

Quasi-bağımsızlık altındaki tahminler için doğrudan bir formül olmadığından, doğrudan olmayan yöntemler tercih edilir. Birçok tamamlanmamış tabloda yapısal sıfırların dışında kalan hücrelerin doğrudan tahmini mümkündür. Aktif olmayan hücreler için iki kural vardır. Bu iki kuralın en önemlisi; quasi bağımsızlık altında aktif olmayan hücreler için EÇOB tahminidir. O hücreler için EÇOB tahmini gözlenen sayılardır ve eğer bu tür hücreler hemen tanımlanabilirse analiz için bu hücreler atılıp, analizden sonra tekrar yerine koyulabilir. İkinci kural ise; EÇOB tahminleri tarafından uyumlu alt tablolardaki tamamlanmamış tabloların bazılarını bozmaktır. Tamamlanmamış tablo için iki kural daha belirtilmiş ki, bu kurallar tablonun iki sınıfı için en çok olabilirliklerinin doğrudan tahminidir. Dört kural tüm tamamlanmamış durumlar için direk tahmini oluşturmada EÇOB tahminine doğrudan yaklaşımı tanımlar. Ayrıştırılmayan tablolar için, (3.16)'da verilen  $\delta_{ij}$ 'lar göz önüne alınarak sözü edilen dört kural aşağıda verilmiştir:

#### 1.KURAL: (Hücre izolasyonu)

Eğer bazı hücreler için  $\delta_{ij}=1$  fakat aynı satır (sütun) da kalan  $\delta_{ij}$  nin tümü sıfır ise

$$\hat{m}_{ij} = x_{ij}$$

dir. Bu durum marjinal toplamın EÇOB tahmini tek olarak belirlemesinden kaynaklanmaktadır.  $(i,j)$  hücreleri aktif olmadığına silinebilir ve tahmine devam edilir. Çizelge 3.2'de verilen 3x3'lük bir tablo için hücre izolasyonu şu şekilde ifade edilmektedir:

**Çizelge 3.2.** 3x 3 biçiminde tamamlanmamış tablo

$m_{11}$	$m_{12}$	$m_{13}$
$m_{21}$	$m_{22}$	-
-	-	$m_{33}$

Çizelge 3.2'deki tablodan üçüncü satır silindiğinde (1,3) hücresi aktif olmayan bir hücredir. (1,3) hücresi, hücre izolasyonunu oluşturur. Bu durumda üçüncü kolon silinebilir. Böylece Çizelge 3.3'de verilen 2x2'lik tablo elde edilir (Bishop *et al.* 1975).

**Çizelge 3.3.** 2x2 biçiminde tamamlanmış tablo

$m_{11}$	$m_{12}$
$m_{21}$	$m_{22}$

Bu durumda, tamamlanmış tablolar için EÇOB tahminleri elde edilebilir.

## **2.KURAL(Yarı-ayrılabilirlik)**

Ayrılamaz tamamlanmamış tablolar; eğer tek satır veya sütunu atarak iki veya daha fazla ayrı alt tablolara ayrılabilirse; yarı ayrılabilir olarak tanımlanır. Örneğin; bir tablo tek kolonu atılarak yarı ayrılabilir tablo oluştursun ve bu tabloda kolon silinmesiyle ayrılabilir alt tabloların birine uygun olarak, her kümesi satırların kümesi olarak parçalansın. Buradan satırların parçalanmış alt kümelerinin her birinde beklenen hücre sayıları tahmin edilebilir. Parçalanmış satırların her kümesi için EÇOB tahmini kesin olarak bulunduğu, tüm tablo için EÇOB tahmini kesin olarak elde edilmiş olur. Çizelge 3.4'de verilen 5 x 5'lik tamamlanmamış tablo göz önüne alınsın.



**Çizelge 3.4** 5x5 biçiminde tamamlanmamış tablo yapısı

$m_{11}$	$m_{12}$	-	-	-
$m_{21}$	$m_{22}$	-	-	$m_{25}$
-	-	$m_{33}$	$m_{34}$	$m_{35}$
-	-	$m_{43}$	$m_{44}$	-
-	-	$m_{53}$	$m_{54}$	$m_{55}$

Bu tamamlanmamış tabloda beşinci kolon silindiğinde iki alt tablo oluşur ki ,bu durum da bu tablo yarı ayrılabilir olur. Tablonun ilk iki satırı ve son 3 satırı ayrıştırılırsa; Çizelge 3.5’de verilen ayrılmış alt tablo elde edilir.

**Çizelge 3.5** Ayrılmış alt tablo yapısı

1			2		
$m_{11}$	$m_{12}$	-	$m_{33}$	$m_{34}$	$m_{35}$
$m_{21}$	$m_{22}$	$m_{25}$	$m_{43}$	$m_{44}$	-
			$m_{53}$	$m_{54}$	$m_{55}$

İlk tabloda  $m_{25}$ ’i içeren alt tablo bir hücre izolasyonudur ve bu değeri içeren sütunun silinmesiyle 2 x 2’lik alt tablo oluşur ve doğrudan EÇOB tahmini hesaplanabilir. Orijinal tablo aktif olmayan bir tablo olmasına rağmen orijinal tablo içinde (2,5) hücresi bir hücre izolasyonu değildir. İkinci alt tabloya aşağıdaki üçüncü ve dördüncü kurallar uygulanarak doğrudan EÇOB tahminleri hesaplanabilir (Bishop and Fienberg 1975).

### 3.KURAL : (Blok-üçgensel tablolar)

Eğer satırların ve sütunların uygun permütasyonlarından sonra

$$\delta_{ij} = 0 \quad \delta_{kl} = 0 \quad \forall k \geq i, l \geq j$$

oluyorsa tamamlanmamış tablolara “blok-üçgensel “ form diyebiliriz.

Blok –üçgensel tablo Çizelge 3.6 ve Çizelge 3.7 ile belirtilmiştir.

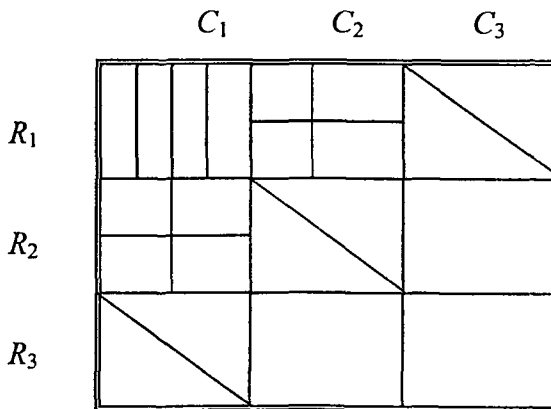
**Çizelge 3.6** Blok-üçgensel tamamlanmamış alt tablo yapısı

1				2			
$m_{11}$	$m_{12}$	$m_{13}$	$m_{14}$	-	-	$m_{13}$	$m_{14}$
$m_{21}$	$m_{22}$	$m_{23}$	-	-	-	$m_{23}$	$m_{24}$
$m_{31}$	$m_{32}$	-	-		$m_{31}$	$m_{32}$	$m_{33}$
$m_{41}$	-	-	-		$m_{41}$	$m_{42}$	$m_{43}$

**Çizelge 3.7** Blok-üçgensel tamamlanmamış alt tablo yapısı

3				
	$m_{11}$	$m_{12}$	$m_{13}$	$m_{14}$
		$m_{22}$	$m_{23}$	$m_{24}$
	-	$m_{32}$	$m_{33}$	$m_{34}$
	-	-	$m_{43}$	$m_{44}$
-	-		$m_{53}$	$m_{54}$

Bu tür tablolara blok-üçgensel tablolar denebilir çünkü satırların ve sütunların uygun permütasyonundan sonra yapısal sıfır olmayan hücreler, üçgenin hipotenüsü boyunca hücrelerin blokları ile sağ-köşeli üçgenlerin bir düzenini oluşturur. Genelde tabloların sağ köşesinin altında yapısal sıfır hücre olan tablolar tercih edilir (Şekil 3.1).



**Şekil 3.1** Blok-üçgensel tabloların genel gösterimi

Şekilde taralı olmayan alanlar yapısal sıfır olmak üzere hücrelerin 2 bloğunda yapısal sıfırlar bulunmaktadır. Yapısal sıfır içermeyen gözelerle ilişkin değerlerin EÇOB tahminleri,

$$\delta_{ij} = 0 \quad i \geq R_1 + 1, j \geq C_1 + C_2 + 1, i \geq R_1 + R_2 + 1, j \geq C_1 + 1$$

olmak üzere köşegendeki bloğa bakılarak;

$$i = R_1 + 1, R_1 + 2, \dots, R_1 + R_2$$

$$j = C_1 + 1, C_1 + 2, \dots, C_1 + C_2$$

için,

$$\begin{aligned} m_{i+} m_{+j} &= \left( \sum_{l=1}^{C_1+C_2} \exp(u + u_{1(i)} + u_{2(l)}) \right) \left( \sum_{k=1}^{R_1+R_2} \exp(u + u_{1(k)} + u_{2(j)}) \right) \\ &= \exp(u + u_{1(i)} + u_{2(j)}) \sum_{l=1}^{C_1+C_2} \sum_{k=1}^{R_1+R_2} \exp(u + u_{1(k)} + u_{2(l)}) \\ &= m_{ij} \left( \sum_{k=1}^{R_1+R_2} \sum_{l=1}^{C_1+C_2} m_{kl} \right) \\ \sum_{k=1}^{R_1+R_2} \sum_{l=1}^{C_1+C_2} m_{kl} &= m_{++} - \sum_{i=R_1+R_2+1}^{R_1+R_2+R_3} m_{i+} - \sum_{j=C_1+C_2+1}^{C_1+C_2+C_3} m_{+j} \end{aligned} \quad (3.22)$$

olarak elde edilir. Burada  $u$ -terimi, tablodaki bütün hücrelerin olasılıklarının logaritmasının genel ortalamasını belirtirken,  $u_{i(j)}$ -terimi de  $i$ -inci değişkenin  $j$ -inci düzeyindeki hücre olasılıklarının logaritmasını ifade etmektedir. Bu blok için  $m_{ij}$ , ilişkin marjinal toplamların doğrudan fonksiyonu gibi yazılabilir

EÇOB tahmin ;

$$\hat{m}_{ij} = \frac{x_{i+} x_{+j}}{\sum_{k=1}^{R_1+R_2} \sum_{l=1}^{C_1+C_2} x_{kl}} \quad i = R_1 + 1, \dots, R_1 + R_2 \quad j = C_1 + 1, \dots, C_1 + C_2 \quad (3.23)$$

dir (Bishop and Fienberg 1975).

#### 4.KURAL : (Blok –basamak tablolar)

Uygun satır ve sütunların permütasyonundan sonra, satırların kümesi tablolara bölünebilir. Sıfır olmayan hücrelerin sırası dikdörtgenlerden birini içerir ve bunun gibi dikdörtgenler kolonlarda hemen diğerinin altında ve aşağısında ise ayrılamaz

tamamlanmamış tabloya blok-basamak tablo denir (Bishop and Fienberg 1975). Böyle bir tablo Şekil 3.2 de verilmiştir.

	$C_1$	$C_2$	$C_3$	$C_4$
$R_1$			/	
$R_2$		/		
$R_3$				
$R_4$		/		

Şekil 3.2 Blok-basamak tabloların genel gösterimi

Şekil 3.2 incelendiğinde taralı olmayan alanlar yapısal sıfır blokları göstermek üzere, tüm blok-basamak tamamlanmamış tablolar, quasi bağımsızlık altında sıfır olmayan gözlemlere ilişkin beklenen sıklıklar için kapalı bir düzende EÇOB tahminlere sahiptir. Blok -basamak tablosunda dikdörtgen sıralarından biri tek bir satır veya sütun olduğunda, o satır veya sütun hücre izolasyonu ya da ayrılabilir tablo olabilir. Tablo için EÇOB tahmin, yedi taralı bloğun her birinin EÇOB tahminlerin toplamıdır ve bu toplam gözlenmiş sayıların toplamına eşittir. Bu yedi bloktan iki tanesi için EÇOB tahmin, taralı hücreler için,

$$i = R_1 + 1, \dots, R_1 + R_2$$

$$j = C_1 + C_2 + 1, \dots, C_1 + C_2 + C_3$$

$$\begin{aligned}
m_{i+}m_{+j} &= \left( \sum_{l=C_1+1}^{C_1+C_2+C_3} \exp(u + u_{1(i)} + u_{2(l)}) \right) \left( \sum_{k=1}^{R_1+R_2} \exp(u + u_{1(k)} + u_{2(j)}) \right) \\
&= \exp(u + u_{1(i)} + u_{2(j)}) \left( \sum_{l=C_1+1}^{C_1+C_2+C_3} \sum_{k=1}^{R_1+R_2} \exp(u + u_{1(k)} + u_{2(l)}) \right) \quad (3.24) \\
m_{ij} &= \frac{m_{i+}m_{+j}}{\sum_{l=C_1+1}^{C_1+C_2+C_3} \sum_{k=1}^{R_1+R_2} \exp(u + u_{1(k)} + u_{2(l)})}
\end{aligned}$$

dir.

Hücrelerin blok içinde tüm  $i, j$  üzerinden toplamı

$$\frac{\sum_{i=R_1+1}^{R_1+R_2} \sum_{j=C_1+C_2+1}^{C_1+C_2+C_3} m_{i+}m_{+j}}{\sum_{i=R_1+1}^{R_1+R_2} \sum_{j=C_1+C_2+1}^{C_1+C_2+C_3} m_{ij}} = \sum_{l=C_1+1}^{C_1+C_2+C_3} \sum_{k=1}^{R_1+R_2} \exp(u + u_{1(k)} + u_{2(l)}) \quad (3.25)$$

dir.

Sonuçta  $m_{ij}$ , satır ve sütunların toplam beklenen sıklıkları,  $\{m_{i+}\}$  ve  $\{m_{+j}\}$ , marjinal toplamaların bir fonksiyonudur. EÇOB tahmini bu hücreler için yazılırsa,

$$\hat{m}_{ij} = \frac{x_{i+}x_{+j} \left( \sum_{k=R_1+1}^{R_1+R_2} \sum_{l=C_1+C_2+1}^{C_1+C_2+C_3} x_{kl} \right)}{\sum_{k=R_1+1}^{R_1+R_2} \sum_{l=C_1+C_2+1}^{C_1+C_2+C_3} x_{k+}x_{+l}} \quad \begin{aligned} i &= R_1 + 1, \dots, R_1 + R_2 \\ j &= C_1 + C_2 + 1, \dots, C_1 + C_2 + C_3 \end{aligned} \quad (3.26)$$

şeklinde belirtilmiştir. Formül oldukça karmaşık görünmesine rağmen, temel yapısı oldukça basittir. İlk olarak tamamlanmış tablolara ilişkin kolon ve satır toplamalarını, bloktaki gözlemlerin toplam sayılarının toplamalarını toplayıp onların marjinal toplamını marjinal toplama bölüp, blok için gözlenmiş toplamla çarpılır (Bishop and Fienberg 1975).

### 3.4. Kapalı Form Tahminleri

Bir çok yazar, tamamlanmamış iki yönlü olumsuzluk tabloların EÇOB tahminlerinin elde edilmesi için yaklaşımlar üretmiştir. İki yönlü olumsuzluk tablolar üç tamamlayıcı kısım halinde ayırarak log-doğrusal modellerin yardımıyla EÇOB tahminlerini elde edilmektedir. Gösterimi kolay ve boyutlar artsa bile log-doğrusal modeller kullanılmaktadır çünkü çoklu tamamlayıcı tablolara paylaştırılarak kolayca

genelleştirme yapılabileceğinden, gözlemlerin Poisson dağılımı gösterdiği düşünülmektedir (Baker *et al.* 1991).

$\{z_{ijkl}\}$  hücre değerlerini göstermek üzere;

$M_I = M_J = 1$  için hücre değerleri  $\{z_{ij11}\}$

$M_I = 1$  ve  $M_J = 2$  de tamamlayıcı  $I$  payına ait hücre değerlerinin toplamı  $\{z_{i+12}\}$ ,

$M_I = 2$  ve  $M_J = 1$  de tamamlayıcı  $J$  payına ait hücre değerlerinin toplamı  $\{z_{+j21}\}$ ,

$M_I = 2$  ve  $M_J = 2$  de tamamlayıcı hücre değerlerinin toplamı  $\{z_{++22}\}$ ,

$N = z_{++++}$  toplam gözlem sayısıdır.

Eksik gözlemlere sahip olumsuzluk tablo yapısı Çizelge 3.7’de belirtilmiştir.

**Çizelge 3.8.** Eksik gözlemlerin bulunduğu 2x2’lik tablo yapısı

		$M_J = 1$		$M_J = 2$
		$J = 1$	$J = 2$	(Eksik gözlem)
$M_I = 1$	$I = 1$	$z_{1111}$	$z_{1211}$	$z_{1+12}(w_{1+})$
	$I = 2$	$z_{2111}$	$z_{2211}$	$z_{2+12}(w_{2+})$
$M_I = 2$ (Eksik gözlem)		$z_{1+21}$ ( $h_{+1}$ )	$z_{+221}$ ( $h_{+2}$ )	$z_{++22}(t_{++})$

### 3.4.1 Log- Doğrusal model yapısı ve tahminler

Tamamlayıcı kısımlardan oluşan olumsuzluk tablolar için log-doğrusal model yapısı, eksik verilerle diğer değişkenleri birleştirir. Log-doğrusal model yapısı,

$$\log(\mu_{ijkl}) = \exp\{u + u_I(i) + u_J(j) + u_{M_I}(k) + u_{M_J}(l) + u_{IJ}(ij) + u_{IM_I}(ik) + u_{IM_J}(il) + u_{JM_I}(jk) + u_{JM_J}(jl) + u_{M_I M_J}(kl)\}$$

biçimindedir.

EÇOB tahminlerinin elde edilmesinde kolaylık olması amacıyla, ifade;

$$\begin{aligned}
 m_{ij} &= \exp \{u + u_I(i) + u_J(j) + u_{M_I}(1) + u_{M_J}(1) + u_{IJ}(ij) + u_{IM_I}(i1) \\
 &\quad + u_{IM_J}(i1) + u_{JM_I}(j1) + u_{JM_J}(j1) + u_{M_I M_J}(11)\} \\
 a_{ij} &= \exp \{ - 2 [u_{M_I}(1) + u_{IM_I}(i1) + u_{JM_I}(j1) + u_{M_I M_J}(11)] \} \\
 b_{ij} &= \exp \{ - 2 [u_{M_J}(1) + u_{IM_J}(i1) + u_{JM_J}(j1) + u_{M_I M_J}(11)] \} \\
 g &= \exp \{ 4 u_{M_I M_J}(11) \}
 \end{aligned}$$

ve

$$\begin{aligned}
 \mu_{ij11} &= m_{ij}, & \mu_{ij12} &= m_{ij} b_{ij}, \\
 \mu_{ij21} &= m_{ij} a_{ij}, & \mu_{ij22} &= m_{ij} a_{ij} b_{ij} g, \\
 m_{ij} \geq 0, & a_{ij} \geq 0, b_{ij} \geq 0, g \geq 0 \text{ ve } \sum_i \sum_j m_{ij} (1 + a_{ij} + b_{ij} + a_{ij} b_{ij} g) = N
 \end{aligned} \tag{3.27}$$

ve

$$\begin{aligned}
 m_{ij} &= N \Pr(I = i, J = j, M_I = 1, M_J = 1) \\
 a_{ij} &= \frac{\Pr(M_I = 2 \text{ ve } M_J = 1 | I = i, J = j)}{\Pr(M_I = 1 \text{ ve } M_J = 1 | I = i, J = j)} \\
 b_{ij} &= \frac{\Pr(M_I = 1 \text{ ve } M_J = 2 | I = i, J = j)}{\Pr(M_I = 1 \text{ ve } M_J = 1 | I = i, J = j)} \\
 g &= \frac{\Pr(M_I = 1 \text{ ve } M_J = 1 | I = i, J = j) \Pr(M_I = 2 \text{ ve } M_J = 2 | I = i, J = j)}{\Pr(M_I = 1 \text{ ve } M_J = 2 | I = i, J = j) \Pr(M_I = 2 \text{ ve } M_J = 1 | I = i, J = j)}
 \end{aligned} \tag{3.28}$$

biçiminde verilmiştir.

Eğer  $g=1$  (odds oranı) ise eksik veri yapısı bağımsızdır.  $I$  ve  $J$  için marjinal olasılıklar;

$$\pi_{ij..} = P(I = i, J = j) = m_{ij} (1 + a_{ij} + b_{ij} + a_{ij} b_{ij} g) / N; \tag{3.29}$$

dır.

$\mathcal{J}$  de veriler gözlendiğinde ve  $\mathcal{I}$  da veriler gözlenemediğinde, eksik verilerin olasılığı;

$$\psi_{ij} = P(M_I = 2 | M_J = 1, I = i, J = j) = a_{ij} / (1 + a_{ij}); \tag{3.30}$$

ve

$\mathcal{I}$  da veriler gözlendiğinde ve  $\mathcal{J}$  deki verilerin gözlenememesi olasılığı;

$$\eta_{ij} = P(M_J = 2 | M_I = 1, I = i, J = j) = b_{ij} / (1 + b_{ij}); \tag{3.31}$$

dir (Baker et al. 1992).

Tamamlanmamış olumsuzluk tablolarının EÇOB tahminlerine ilişkin parametre tahminleri farklı yapılara göre verilmiştir. Model yapılarında kullanılan  $\alpha_{..}, \alpha_i, \alpha_j$  terimleri  $a_{ij}$  terimine ve  $\beta_{..}, \beta_i, \beta_j$  terimleri de  $b_{ij}$  terimine eşittir. ‘.’ İfadesi parametrenin uygun indisin sabit olduğunu belirtmektedir. Bu model yapıları:

**a) Her değişkende eksik veri sabit { EM algoritması için basit bir yaklaşım} olduğunda,**

Model  $(\alpha_{..}, \beta_{..})$  s.d.=  $n_1+n_2-1$

$$\hat{m}_{ij}^{(0)} = z_{ij11}$$

$$\hat{m}_{ij}^{(t+1)} = \frac{\left[ z_{ij11} + z_{i+12} \frac{\hat{m}_{ij}^{(t)}}{\hat{m}_{i+}} + z_{j21} \frac{\hat{m}_{ij}^{(t)}}{\hat{m}_{+j}} \right] z_{++11}}{z_{++11} + z_{++12} + z_{++21}}$$

$$\hat{\alpha}_{..} = \frac{z_{++21}}{z_{++11}} \quad (3.32)$$

$$\hat{\beta}_{..} = \frac{z_{++12}}{z_{++11}}$$

$$\hat{g} = \frac{z_{++11} z_{++22}}{z_{++12} z_{++21}}$$

biçimindedir.

**b) Bir değişkende eksik veri sabit ve diğer değişkende eksik veri bir öncekinin anlamına bağlı olduğunda,**

Model  $(\alpha_{..}, \beta_i)$  s.d.=  $n_1-1$

$$\hat{m}_{ij} = \frac{z_{ij11} z_{+j+1} z_{++11}}{z_{+j11} z_{++11}}$$

$$\hat{\alpha}_{..} = \frac{z_{++21}}{z_{++11}}$$

$$\hat{\beta}_i = \frac{z_{i+12}}{\hat{m}_{i+}}$$

$$\hat{g} = \frac{z_{++11} z_{++22}}{z_{++12} z_{++21}}$$

biçimindedir.

Model  $(\alpha_j, \beta_{..})$  s.d.=  $n_2-1$

$$\hat{m}_{ij} = \frac{z_{ij11} z_{i+1+} z_{++11}}{z_{i+11} z_{++1+}}$$

$$\hat{\alpha}_j = \frac{z_{+j21}}{\hat{m}_{+j}}$$

$$\hat{\beta}_{..} = \frac{z_{++12}}{z_{++11}}$$

$$\hat{g} = \frac{z_{++11} z_{++22}}{z_{++12} z_{++21}}$$

(3.33)



c) Bir değişkende eksik veri sabit ve diğer değişkendeki eksik veri kendisinin anlamına bağlı olduğunda,

Model  $(\alpha_{..}, \beta_{.j})$  s.d.=  $n_1-1$

Model  $(\alpha_{i.}, \beta_{..})$  s.d.=  $n_2-1$

$$\hat{m}_{ij} = \frac{z_{ij11} z_{+j+1} z_{++11}}{z_{+j11} z_{++++}}$$

$$\hat{m}_{ij} = \frac{z_{ij11} z_{i+1+} z_{++11}}{z_{i+11} z_{++++}}$$

$$\hat{\alpha}_{..} = \frac{z_{++21}}{z_{++11}}$$

$$\hat{\alpha}_{i.}; \sum_i \hat{m}_{ij} \hat{\alpha}_{i.} = z_{+j21}$$

$$\hat{\beta}_{.j}; \sum_i \hat{m}_{ij} \hat{\alpha}_{i.} = z_{i+12}$$

$$\hat{\beta}_{..} = \frac{z_{++12}}{z_{++11}}$$

$$\hat{g} = \frac{z_{++11} z_{++22}}{z_{++12} z_{++21}}$$

$$\hat{g} = \frac{z_{++11} z_{++22}}{z_{++12} z_{++21}}$$

(3.34)

biçimindedir. Bu yaklaşım  $n_1=n_2$  ve negatif olmayan tahminleri gerektirir.

d) Her değişkendeki eksik veri aynı değişkenin anlamına bağlı olduğunda,

Model  $(\alpha_{i.}, \beta_{.j})$  s.d.=  $n_1-1$

Model  $(\alpha_{.j}, \beta_{.j})$  s.d.=  $n_2-1$

$$\hat{m}_{ij} = z_{ij11}$$

$$\hat{m}_{ij} = z_{ij11}$$

$$\hat{\alpha}_{i.}; \sum_i z_{ij11} \hat{\alpha}_{i.} = z_{+j21}$$

$$\hat{\alpha}_{.j} = \frac{z_{+j21}}{z_{+j11}}$$

$$\hat{\beta}_{.j} = \frac{z_{i+12}}{z_{i+11}}$$

$$\hat{\beta}_{.j}; \sum_j z_{ij11} \hat{\beta}_{.j} = z_{i+12}$$

$$\hat{g} = \frac{z_{++22}}{\sum_i \hat{m}_{i+} \hat{\alpha}_{i.} \hat{\beta}_{.j}}$$

$$\hat{g} = \frac{z_{++22}}{\sum_j \hat{m}_{+j} \hat{\alpha}_{.j} \hat{\beta}_{.j}}$$

(3.35)

biçimindedir. Bu yaklaşım  $n_1=n_2$  ve negatif olmayan tahminleri gerektirir.

e) Her değişkendeki eksik veri anlama bağlı olduğunda,

Model  $(\alpha_{i.}, \beta_{.j})$  s.d.= 0

$$\hat{m}_{ij} = z_{ij11}$$

$$\hat{\alpha}_{i.}; \sum_i z_{ij11} \hat{\alpha}_{i.} = z_{+j21}$$

$$\hat{\beta}_{.j}; \sum_j z_{ij11} \hat{\beta}_{.j} = z_{i+12}$$

$$\hat{g} = \frac{z_{++22}}{\sum_j \hat{m}_{ij} \hat{\alpha}_{i.} \hat{\beta}_{.j}}$$

(3.36)

biçimindedir. Bu yaklaşım  $n_1=n_2$  ve negatif olmayan tahminleri gerektirir.

f) Her değişkendeki eksik veri diğer değişkenin anlamına bağlı olduğunda,

Model  $(\alpha_j, \beta_i)$  s.d.= 0

$$\begin{aligned}\hat{m}_{ij} &= z_{ij11} \\ \hat{\alpha}_j &= \frac{z_{+j21}}{z_{+j11}} \\ \hat{\beta}_i &= \frac{z_{i+12}}{z_{i+11}} \\ \hat{g} &= \frac{z_{++22}}{\sum_j \hat{m}_{ij} \hat{\alpha}_j \hat{\beta}_i}\end{aligned}\tag{3.37}$$

biçimindedir (Baker *et al.* 1992).

### 3.5 Tamamlanmamış İki Yönlü Olumsuzluk Tablolarda Uyum İyiliği Testi

Tamamlanmamış olumsuzluk tablolara ilişkin uyum iyiliği testi, tamamlanmış olumsuzluk tabloların uyum iyiliği testi ile benzerlik göstermektedir. Belli modeller için EÇOB tahmininin temelindeki modelin uyum iyiliğinin testi ya Pearson ya da olabilirlik oran Ki-kare testi ile hesaplanır (Baker *et al.* 1992).

Olabilirlik oran istatistiği;

$$G^2 = -2 \left\{ \sum_i \sum_j z_{ij11} \log \left[ \frac{\hat{m}_{ij}}{z_{ij11}} \right] + \sum_i z_{i+12} \log \left[ \frac{\sum_j \hat{m}_{ij} \hat{\beta}_j}{z_{i+12}} \right] + \sum_j z_{+j21} \log \left[ \frac{\sum_i \hat{m}_{ij} \hat{\alpha}_i}{z_{+j21}} \right] + z_{++22} \log \left[ \frac{\sum_i \sum_j \hat{m}_{ij} \hat{\alpha}_i \hat{\beta}_j \hat{g}}{z_{++22}} \right] \right\}$$

biçimindedir.

Uyum iyiliği testi için olabilirlik oran istatistiğindeki gerekli parametrelerin tahminleri , log-doğrusal model yapısı altında farklı model yapılarına göre verilmiştir. Bu modeller EÇOB tahminleri ve uyum iyiliği testi için gerekli parametrelerin tahminlerini içermektedir. Modeller ve modellere göre tahminler (3.32) – (3.37) ifadeleri ile verilmiştir (Baker *et al.* 1991).

#### 4. UYGULAMA

Bu bölümde üçüncü bölümde verilen tamamlanmamış olumsuzluk tablolar simülasyonla elde edilecek ve bu tablolara ilişkin değerlendirmeler daha önceki bölümde verilen ifadelerle ilgili kalınarak yapılacaktır. Çalışma verileri, simülasyonla Poisson ( $\lambda=0.5$ ) dağılımına sahip kitleden 100 ve 40 birimlik örneklem alınarak MATHLAB 6.05 paket programı ile üretilecektir. Öncelikle verilerin yapısı belirtilecek ve verilere ilişkin oluşturulan çapraz tablolarda SPSS 11.5 paket programı kullanılacaktır. Tablolarda gözlemlere ait En çok olabilirlik tahminleri elde etmek için örneklem model yöntemlerinden Poisson model yapısı verilecektir. Poisson model yapısının verilmesinin nedeni, dağılımının kolay elde edilebilmesi olarak gösterilebilir. En çok olabilirlik tahminleri kapalı-form tahmin yöntemleri kullanılarak uygun model yapısına göre tahminler ve olabilirlik oran istatistiğinin değeri elde edilerek, bu değere göre değişkenlerin bağımsız olup olmadığına karar verilecektir. Ayrıca en çok olabilirlik tahminler SAS ve STATA 8.0 paket programları kullanılarak da elde edilecektir. Kapalı form tahmin yöntemi kullanılarak elde edilen sonuçlar, eksik veriler göz önüne alınmadan elde edilen sonuçlarla karşılaştırılacaktır.

Üretilen veriler; kesikli kategorik veri şeklindedir. Verilere ilişkin olumsuzluk tablo Çizelge 4.1'de verilmiştir.

**Çizelge 4.1** 100 birimlik örnek için eksik gözlemleri içeren tamamlanmamış olumsuzluk tablo

Y	X			Toplam
	1	2	Eksik	
1	14	8	4	26
2	36	21	4	61
Eksik	6	3	4	13
Toplam	56	32	12	100

Çalışmadaki istatistiksel değerlendirmeler için kullanılacak veriler MATHLAB 6.05 paket programında üretilmiştir. Program yapısı aşağıdaki şekildedir.

```
Lambda=0.5  
Sample=poissrnd(lambda,1,100)
```

Çapraz tablolarının elde edilmesinde SPSS for Windows 11.5 kullanılmıştır. Verilere ilişkin parametre tahminleri için SAS version 8.0 paket programı kullanılmıştır. Ayrıca STATA 8.0 paket programına göre de parametre tahminleri verilmiştir. Bu çalışmada SAS GENMOD işlemi kullanılarak olasılıkların EÇOB tahminleri elde edilmiştir. SAS ve STATA paket programlarına ilişkin tahminler EK 1 ve EK 2 de verilmiştir.

SAS komutları içinde;

Dist: Model içerisinde kullanılan olasılık dağılımının belirtildiği komuttur.

Link: Model içerisinde kullanılan bağlantı fonksiyonunu tanımlar. Poisson dağılımı için bağlantı fonksiyonu burada identity olarak alınmıştır.

Bu komutlar STATA 8.0 programı içinde aynıdır.

SAS programı sonucunda elde edilen değerler ;

$$p_{11} = 0.1913$$

$$p_{12} = 0.1076$$

$$p_{21} = 0.4452$$

$$p_{22} = 1 - (0.1913 + 0.1076 + 0.4452) = 0.2559$$

ve

STATA programı için elde edilen değerler;

$$p_{11} = 0.18036$$

$$p_{12} = 0.0985$$

$$p_{21} = 0.4352$$

$$p_{22} = 0.28594$$

olarak elde edilmiştir.

Tahminlere ait Poisson model yapısı için tasarım matris yapısı,

$$E \begin{bmatrix} z_{11} \\ z_{12} \\ z_{21} \\ z_{22} \\ w_{1+} \\ w_{2+} \\ h_{+1} \\ h_{+2} \end{bmatrix} = \begin{bmatrix} 79 & 0 & 0 \\ 0 & 79 & 0 \\ 0 & 0 & 79 \\ -79 & -79 & -79 \\ 8 & 8 & 0 \\ -8 & -8 & 0 \\ 9 & 0 & 9 \\ -9 & 0 & -9 \end{bmatrix} \begin{bmatrix} 0,1913 \\ 0,1076 \\ 0,4452 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 79 \\ 0 \\ 8 \\ 0 \\ 9 \end{bmatrix}$$

ve EÇOB tahmin değerleri,

$$\begin{aligned} E(z_{11}) &= 13.351 & E(w_{1+}) &= 2.3592 \\ E(z_{12}) &= 9.9461 & E(w_{2+}) &= 5.6408 \\ E(z_{21}) &= 31.3314 & E(h_{+1}) &= 5.0904 \\ E(z_{22}) &= 24.3715 & E(h_{+2}) &= 3.9096 \end{aligned}$$

biçiminde elde edilmiştir.

Kapalı-form yapısı için Çizelge 4.1 kullanılarak elde edilen EÇOB tahmin değerleri ise,

$$\begin{aligned} \hat{m}_{11} &= \frac{14 \times 26 \times 79}{22 \times 87} = 15.024 & \hat{\alpha}_{.1} &= \frac{6}{50.008} \\ \hat{m}_{12} &= \frac{8 \times 26 \times 79}{22 \times 87} = 8.585 & \hat{\alpha}_{.2} &= \frac{3}{28.992} \\ \hat{m}_{21} &= \frac{36 \times 61 \times 79}{57 \times 87} = 34.984 & \hat{\beta}_{..} &= \frac{8}{79} \\ \hat{m}_{22} &= \frac{21 \times 61 \times 79}{57 \times 87} = 20.407 & \hat{g} &= \frac{79 \times 4}{8 \times 9} = 4.389 \end{aligned}$$

biçiminde elde edilmiştir.

Verilere ilişkin uyum iyiliği testi için model yapısı; Model  $(\alpha_j, \beta_{..})$  olarak alınırsa, Model  $(\alpha_j, \beta_{..})$  yapısı;  $X$  değişkenindeki eksik verilerin  $Y$  değişkenindeki eksik verilere bağlı olduğunu vurgulamaktadır. Bir değişkenindeki eksiklik durumu sabit iken, diğer değişkenindeki eksiklik bir önceki değişkene bağlıdır. Bu bağımlılık; bazı değişkenlere ilişkin değerlerin kaydedilmeden izlemeden kaybolması ile açıklanabilir. Bu modelde  $g$  değerine ilişkin güven aralığı  $[0.054; 354.153]$  ve  $\hat{g} = 4.39$  olduğundan her iki değişkenindeki eksik verilerin bağımsız olduğu söylenebilir.

Görüldüğü gibi farklı yöntemlerle elde edilen EÇOB tahmin yöntemleri arasında bir miktar farklılık görülmektedir. Elle hesaplama daha kolay olduğundan Kapalı-form tahminlerine göre olabilirlik oran istatistiğinin değeri elde edilecektir. Bu tahmin değerlerine bağlı olarak olabilirlik oran istatistiğinin değeri;

$$G^2 = -2 \left\{ \begin{array}{l} 14 \times \log\left(\frac{15.024}{14}\right) + 8 \times \log\left(\frac{8.585}{8}\right) + 36 \times \log\left(\frac{34.984}{36}\right) + 21 \times \log\left(\frac{20.407}{21}\right) \\ + 4 \times \log\left(\frac{15.024 \times \left(\frac{8}{79}\right) + 8.585 \times \left(\frac{8}{79}\right)}{4}\right) + 4 \times \log\left(\frac{34.984 \times \left(\frac{8}{79}\right) + 20.407 \times \left(\frac{8}{79}\right)}{4}\right) \\ + 6 \times \log\left(\frac{15.024 \times \left(\frac{6}{50.008}\right) + 34.984 \times \left(\frac{6}{50.008}\right)}{6}\right) + 3 \times \log\left(\frac{8.585 \times \left(\frac{3}{28.992}\right) + 20.407 \times \left(\frac{3}{28.992}\right)}{3}\right) \\ + 4 \times \log\left(\frac{15.024 \times \frac{6}{50.008} \times \frac{8}{79} \times \frac{316}{72} + 8.585 \times \frac{3}{28.992} \times \frac{8}{79} \times \frac{316}{72} + 34.984 \times \frac{6}{50.008} \times \frac{8}{79} \times \frac{316}{72} + 20.407 \times \frac{3}{28.992} \times \frac{8}{79} \times \frac{316}{72}}{4}\right) \end{array} \right.$$

$$G^2 = 1.571$$

olarak bulunmuştur. 100 birimlik örneklemden üretilen veriler için olabilirlik oran istatistiklerine göre bu iki değişkenin ( $G^2 = 1.571 < \chi^2_{(1;0,05)} = 3.841$ ) olduğu için bağımsız olduğu söylenebilir Eksik gözlemler dikkate alınmasaydı Çizelge 4.2 de verilen tamamlanmış tablo elde edilir. Bu tabloya göre elde edilen uyum iyiliği test sonuçları Çizelge 4.3'de verilmiştir.

**Çizelge 4.2.** . 100 birimlik örnek için eksik gözlemleri içermeyen tamamlanmış olumsuzluk tablo

Y	X		Toplam
	1	2	
1	36	21	57
2	14	8	22
Toplam	50	29	79

**Çizelge 4.3.** 100 birimlik örnek için tamamlanmış tabloya ilişkin uyum iyiliği testi sonuçları

	değer	sd	asimptotik anlamlılık
Pearson Ki-kare	,002	1	,968
olabilirlik oran	,002	1	,968
toplam	79		

Eksik gözlemler dikkate alınmadan elde edilen tamamlanmış tabloya göre iki değişkenin bağımsızlık test sonucunda da, Çizelge 4.3 de görüldüğü üzere değişkenlerin ( $G^2 = 0.002 < \chi^2_{(1;0,05)} = 3.841$ ) olduğu için bağımsız olduğu sonucu elde edilmiştir

40 birimlik örneklem için olumsuzluk tablo aşağıdaki gibidir:

**Çizelge 4.4** 40 birimlik örnek için eksik gözlemleri içeren tamamlanmamış olumsuzluk tablo

Y	X			Toplam
	1	2	Eksik	
1	11	9	1	15
2	7	5	3	21
Eksik	2	1	1	4
Toplam	20	15	4	40

Çizelge 4.4'deki veriler kullanılarak Poisson model yapısına göre EÇOB tahminleri, aşağıda verilmiştir:

$$\begin{aligned}
 E(z_{11}) &= 7.2829 & E(w_{1+}) &= 1.6988 \\
 E(z_{12}) &= 6.2848 & E(w_{2+}) &= 2.3032 \\
 E(z_{21}) &= 9.0656 & E(h_{+1}) &= 1.5333 \\
 E(z_{22}) &= 9.36 & E(h_{+2}) &= 1.4667
 \end{aligned}$$

Model ( $\alpha_j, \beta_{..}$ ) yapısı için 40 birimlik örneklem verileri kullanarak elde edilen EÇOB tahminleri ve olabilirlik oran istatistiği aşağıdaki gibidir:

$$\begin{aligned}
 \hat{m}_{11} &= \frac{7 \times 15 \times 32}{12 \times 36} = 7.2896 & \hat{\alpha}_{.1} &= \frac{2}{18.045} = 0.1108 \\
 \hat{m}_{12} &= \frac{5 \times 15 \times 32}{12 \times 36} = 5.556 & \hat{\alpha}_{.2} &= \frac{1}{13.956} = 0.072 \\
 \hat{m}_{21} &= \frac{11 \times 21 \times 32}{20 \times 36} = 10.267 & \hat{\beta}_{..} &= \frac{4}{32} = 0.125 \\
 \hat{m}_{22} &= \frac{9 \times 21 \times 32}{20 \times 36} = 8.4 & \hat{g} &= \frac{8}{3} = 2.67
 \end{aligned}$$

Bu modelde  $g$  değerine ilişkin güven aralığı  $[0.06; 122.159]$  ve  $\hat{g} = 2.67$  olduğundan her iki değişkendeki eksik verilerin bağımsız olduğu söylenebilir.

Bu tahmin değerlerine bağlı olarak olabilirlik oran istatistiğinin değeri:

$$G^2 = -2 \left\{ \begin{array}{l} 7 \times \log\left(\frac{7.778}{7}\right) + 5 \times \log\left(\frac{5.556}{5}\right) + 11 \times \log\left(\frac{10.267}{11}\right) + 9 \times \log\left(\frac{8.4}{9}\right) \\ + 3 \times \log\left(\frac{7.778 \times (\frac{4}{32}) + 5.556 \times (\frac{4}{32})}{3}\right) + 1 \times \log\left(\frac{10.267 \times (\frac{4}{32}) + 8.4 \times (\frac{4}{32})}{1}\right) \\ + 2 \times \log\left(\frac{7.778 \times (\frac{2}{18.045}) + 10.267 \times (\frac{2}{18.045})}{2}\right) + 1 \times \log\left(\frac{5.556 \times (\frac{1}{13.956}) + 8.4 \times (\frac{1}{13.956})}{1}\right) \\ + 1 \times \log\left(\frac{7.778 \times \frac{2}{18.045} \times \frac{4}{32} \times \frac{8}{3} + 5.556 \times \frac{1}{13.956} \times \frac{4}{32} \times \frac{8}{3} + 10.267 \times \frac{2}{18.045} \times \frac{4}{32} \times \frac{8}{3} + 8.4 \times \frac{1}{13.956} \times \frac{4}{32} \times \frac{8}{3}}{4}\right) \end{array} \right.$$

$$G^2 = 2.061$$

olarak bulunmuştur. 40 birimlik örneklemden üretilen veriler için olabilirlik oran istatistiklerine göre bu iki değişkenin bağımsız olduğu söylenebilir ( $G^2 = 2.061 < \chi^2_{(1;0.05)} = 3.841$ ). Burada da eksik gözlemler dikkate alınmadığında Çizelge 4.5'de verilen tamamlanmış tablo elde edilir. Bu tabloya göre elde edilen uyum iyiliği test sonuçları Çizelge 4.6'da verilmiştir.

**Çizelge 4.5** 40 birimlik örnek için eksik gözlemleri içermeyen tamamlanmış olumsuzluk tablo

Y	X		Toplam
	1	2	
1	11	9	20
2	7	5	12
Toplam	18	14	32

**Çizelge 4.6** 40 birimlik örnek için tamamlanmış tabloya ilişkin uyum iyiliği testi sonuçları

	değer	sd	asimptotik anlamlılık
Pearson Ki-kare	,034	1	,854
olabilirlik oran	,034	1	,854
toplam	32		

Eksik gözlemler dikkate alınmadan elde edilen tamamlanmış tabloya göre iki değişkenin bağımsızlık test sonucunda da, değişkenlerin ( $G^2 = 0.034 < \chi^2_{(1;0.05)} = 3.841$ ) olduğu için bağımsız olduğu sonucu elde edilmiştir



Eksik gözlemler dikkate alındığında  $G^2$  istatistiğinin değerinin arttığı görülmektedir. Daha genel sonuçlar elde edebilmek için farklı örneklem genişliklerinde ve farklı Poisson modelleri üzerinde çalışılabilir. Bunun için daha detaylı simülasyon çalışması yapılabilir.



## 5. TARTIŞMA ve SONUÇ

Tamamlanmamış iki yönlü olumsuzluk tabloların EÇOB tahminleri için yapılan çalışmada iki değişkenin birbirleri ile ilişkisi incelenmiştir. Bu çalışmada eksik gözlemlere ilişkin EÇOB tahminler elde edildikten sonra yapılan uyum iyiliği testi; iki değişken arasında bir ilişki olmadığını göstermektedir. Yine eksik gözlemlerin olmadığı sadece tamamlanmış tablodan elde edilen uyum iyiliği testine göre ise iki değişken arasında bir ilişki olmadığı görülmektedir. Bu durum farklı örneklem alınarak daha detaylı olarak incelenebilir. Ancak analizler sırasında eksik verilerin atılmadan işlemlere dahil edilmesi daha güvenilir sonuçlar ortaya koyacaktır.

Tamamlanmamış tablolarda gözlemlere ait EÇOB tahminler için bazı yöntemler kullanılmaktadır. Bu yöntemlerden elde edilen sonuçlar farklılık göstermektedir. SAS ve STATA paket programından elde edilen tahmin değerleri ile kapalı-form tahminleri altında hesaplanan tahmin değerleri arasında farklılıklar uygulama kısmında belirtilmiştir. Paket programlarından elde edilen sonuçlardaki farklılıklar EK bölümünde gösterilmiştir. Bu farklılık, tahminler elde edildikten sonra yapılan analiz sonuçlarını da etkilemektedir. Bu durumda bilgisayarda paket programlarda elde edilen sonuçlardan daha çok kapalı form tahmin yöntemleri hesaplama kolaylığından dolayı tercih edilir. Farklı örneklem genişliğine, örneklem yöntemlerine ve tablo yapılarına göre ilerde daha detaylı çalışmalar yapmak mümkündür.

EÇOB tahmin yöntemlerinden biri olan Quasi-bağımsızlık altında EÇOB tahmin yöntemi için olumsuzluk tablolarının karesel tablolar olması gerektiği vurgulanmıştır. Bu nedenle kare olmayan olumsuzluk tablolarda, Quasi-bağımsızlık altında EÇOB tahminler elde edilememektedir. Ayrıca tamamlanmamış iki yönlü olumsuzluk tablolarının EÇOB tahminlerinin elde edilmesi için örneklem yöntemlerinin önemi büyüktür. Bu nedenle örneklem yöntemi için dağılımın ve model yapısının belirlenmesi gerekmektedir.

## KAYNAKLAR

- Agresti, A. 1990. "Categorical Data Analysis", A Wiley-Interscience Publication, 1-558, New York
- Baker, S.G. Rosenberger, W.F. and Dersimonian, R. 1992 "Closed Form Estimates for Missing Counts in Two-Way Contingency Tables", *Statistics in Medicine*, 11; 643-657
- Bishop, Y.M.M. and Fienberg, S.E. 1975. "Discrete Multivariate Analysis: Theory and Practice", The MIT Press Cambridge, 1-557, England
- Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. "Maximum Likelihood Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38
- Fuchs, C. 1982. "Maximum Likelihood Estimation and Model Selection in Contingency Tables With Missing Data", *Journal of American Statistical Association*, 77; 270-277
- Lipsitz, S.R. Parzen, M. and Molenberghs, G. 1998. "Obtaining the Maximum Likelihood Estimates in Incomplete RxC Contingency Tables Using a Poisson Generalized Linear Model", *American Statistical Association and Journal of Computational and Graphical Statistics*, 7, 3; 356-376
- McLachlan, G.J. and Krishnan, T. 1997. "The EM Algorithm and Extensions", *Wiley Series in Probability and Statistics*-Wiley & Sons, 1-45
- Ming-Yi, H. 1999. "Model Checking for Incomplete High Dimensional Categorical Data", Copyright by Ming-Yi Hu University of California, 9-28, Los Angeles
- Rapallo, F. 2001. "Markov bases for contingency tables and MCMC using commutative algebra" <http://www.dima.unige.it/~rapallo/mypubl.htm> Erişim tarihi: 15.08.2005
- Rapallo, F. 2003. "Log-Linear Models and Toric Ideals", Italy, 1-100
- Roussas, G.G. 1972. "A First Course in Mathematical Statistics", Addison-Wesley Publishing Company, 242-243

## **EKLER**

**EK 1 TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDA PARAMETRE  
TAHMİNLERİNİN SAS KOMUTLARI ve ÇIKTILARI**

**EK 2 TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDA PARAMETRE  
TAHMİNLERİNİN STATA ÇIKTILARI**



**EK 1 TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDA PARAMETRE  
TAHMİNLERİNİN SAS KOMUTLARI ve ÇIKTILARI**

```

data tez;
input count p11 p12 p21 off;
cards;
14 79 0 0 0
8 0 79 0 0
36 0 0 79 0
21 -79 -79 -79 79
4 8 8 0 0
4 -8 -8 0 8
6 9 0 9 0
3 -9 0 -9 9
;
proc genmod data=tez;
model count =p11 p12 p21 /
dist =poi link=id offset=off noint;
run;

```

The GENMOD Procedure

Model Information

Data Set	WORK.TEZ
Distribution	Poisson
Link Function	Identity
Dependent Variable	count
Offset Variable	off
Observations Used	8

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5	1.6110	0.3222
Scaled Deviance	5	1.6110	0.3222
Pearson Chi-Square	5	1.7407	0.3481
Scaled Pearson X2	5	1.7407	0.3481
Log Likelihood		174.8552	

Algorithm converged.

**EK 1 TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDA PARAMETRE  
TAHMİNLERİNİN SAS KOMUTLARI ve ÇIKTILARI (devamı)**

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi- Square	Pr > ChiSq
Intercept	0	0.0000	0.0000	0.0000	0.0000	.	.
p11	1	<b>0.1913</b>	0.0436	0.1059	0.2767	19.26	<.0001
p12	1	<b>0.1076</b>	0.0348	0.0395	0.1757	9.58	0.0020
p21	1	<b>0.4452</b>	0.0533	0.3407	0.5496	69.79	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Lagrange Multiplier Statistics

Parameter	Chi-Square	Pr > ChiSq
Intercept	0.0578	0.8101

**N=40 SAS paket programı sonuçları**

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi- Square	Pr > ChiSq
Intercept	0	0.0000	0.0000	0.0000	0.0000	.	.
p11	0	<b>0.1690</b>	0.0000	0.1690	0.1690	.	.
p12	0	<b>0.1259</b>	0.0000	0.1259	0.1259	.	.
p21	0	<b>0.3966</b>	0.0000	0.3966	0.3966	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000		

**EK 2 TAMAMLANMAMIŞ OLUMSALLIK TABLOLARDA PARAMETRE  
TAHMİNLERİNİN STATA ÇIKTILARI**

glm count p11 p12 p21 off, family(poisson) link(identity)

Iteration 0: log likelihood = -16.80125  
 Iteration 1: log likelihood = -16.703921  
 Iteration 2: log likelihood = -16.70365  
 Iteration 3: log likelihood = -16.70365

Generalized linear models No. of obs = 8  
 Optimization : ML: Newton-Raphson Residual df = 3  
 Scale parameter = 1  
 Deviance = 1.48322085 (1/df) Deviance = .4944069  
 Pearson = 1.478846826 (1/df) Pearson = .4929489

Variance function:  $V(u) = u$  [Poisson]  
 Link function :  $g(u) = u$  [Identity]  
 Standard errors : OIM

Log likelihood = -16.7036498 AIC = 5.425912  
 BIC = -4.755103776

count	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
p11	<b>.1803672</b>	.0558225	3.23	0.001	.0709572	.2897772
p12	<b>.0985299</b>	.0435774	2.26	0.024	.0131196	.1839401
p21	<b>.4352033</b>	.0752601	5.78	0.000	.2876962	.5827104
off	<b>.9610423</b>	.148694	6.46	0.000	.6696075	1.252477
_cons	<b>.4674921</b>	1.341901	0.35	0.728	-2.162586	3.09757

## ÖZGEÇMİŞ

Adı Soyadı: Anıl AKTAŞ

Doğum Yeri:Ankara

Doğum Tarihi:12 Ocak 1980 yılında

Medeni Hali: Bekar

Yabancı Dili: İngilizce

### Eğitim Durumu(Kurum ve Yıl)

Lise : Dr. Şerafettin Tombulođlu Lisesi 1993-1997

Lisans : Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü 1998-2002

Yüksek Lisans: Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü 2003-2005.