KAYIP GÖZLEM İÇEREN OLUMSALLIK TABLOLARINDA LOG-DOĞRUSAL MODELLER

LOG-LINEAR MODELS FOR CONTINGENCY TABLES WITH MISSING DATA

EMİNE ÖÇAL

DOÇ. DR. AYFER EZGİ YILMAZ ÇAKIROĞLU
Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim- Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2024

ÖZET

KAYIP GÖZLEM İÇEREN OLUMSALLIK TABLOLARINDA LOG-DOĞRUSAL MODELLER

Emine ÖÇAL

Yüksek Lisans, İstatistik Bölümü
Tez Danışmanı: Ayfer Ezgi YILMAZ ÇAKIROĞLU
Eylül 2024, 72 sayfa

Kayıp gözlem, veri setinde bir ya da birden fazla hücrede değer olmaması durumudur. Kayıp gözlemlere sağlık, ekoloji, ekonomi, sosyal bilimler gibi veri elde etmenin zor olduğu alanlarda sıklıkla rastlanmaktadır. Veride kayıp gözlemler olması durumunda veri analizinde gerçeğe uygun olmayan sonuçlar ortaya çıkabilir. Bu nedenle veri analizinden önce kayıp gözlem probleminin çözülmesi gerekir.

Bu tez çalışması kapsamında kayıp gözlem çeşitlerini ve kayıp gözlemle başa çıkma yöntemlerini incelemek, olumsallık tabloları için kullanılan yöntemleri tanıtmak ve logaritmik doğrusal modelleri tanıtarak bu yöntemleri uygulama üzerinden tartışmak amaçlanmaktadır.

Anahtar Kelimeler: Kayıp gözlem, Olumsallık tabloları, Log-doğrusal modeller, Rastgele olmayan kayıp gözlem, Rastgele kayıp gözlem, Tamamen rastgele kayıp gözlem

ABSTRACT

LOG-LINEAR MODELS FOR CONTINGENCY TABLES WITH MISSING DATA

Emine ÖÇAL

Master of Science, Department of Statistics
Supervisor: Ayfer Ezgi YILMAZ ÇAKIROĞLU

September 2024, 72 pages

Missing value is the absence of a value in one or more cells in the data set. Missing values are frequently encountered in areas such as health, ecology, economics, and social sciences where it is difficult to obtain data. If there are missing values in the data, incorrect results may emerge in the statistical analysis. For this reason, the missing value problem should be solved before data analysis.

Within the scope of this thesis, it is aimed to examine the types of missing values and the methods of dealing with missing values, to introduce the methods used for contingency tables and to introduce log-linear models and discuss these methods through application.

Keywords: Missing value, Contingency tables, Log-linear models, Not missing at random, Missing at random, Missing completely at random

TEŞEKKÜR

Yüksek lisans tez çalışmamın her aşamasında, yoğun bir özveri ve sabırla benimle ilgilenen; kısa bir sürede bu çalışmayı ortaya çıkarabilmem için kişisel zamanlarından fedakârlık ederek bana eşlik eden, bilgisiyle yolumu aydınlatan; çalışma süresince, hem akademik hem de özel hayatımda yaşadığım zorluklar karşısında kendimi tükenmiş, yorgun ve çaresiz hissettiğim anlarda bile çalışmaya devam edebilmem için elimden tutup kaldıran; yılmadan, usanmadan beni motive eden ve desteğini bir an olsun eksik etmeyen sevgili danışman hocam Doç. Dr. Ayfer Ezgi Yılmaz'a tüm emekleri için büyük bir minnetle teşekkürlerimi sunarım.

Üniversite eğitimimin en başından itibaren akademik kariyer hedefimde ilham kaynağım olan; enerjik kişiliği, akademik ve entelektüel birikimiyle örnek aldığım değerli hocam Prof. Dr. Gül Ergün'e, hem akademik hem de kişisel olarak donanımlı bir birey olmam için bana her zaman motivasyon veren ve destekleyen sevgili hocam Dr. Öğr. Üyesi Ceren Eda Can'a ve lisans eğitiminden bu yana bir parçası olmaktan gurur duyduğum ve derin bir bağlılıkla yuvam olarak gördüğüm Hacettepe Üniversitesi İstatistik Bölümü'ndeki üzerimde emeği olan tüm hocalarıma içtenlikle teşekkür ederim.

Hayatımın her aşamasında yanımda olan ve beni destekleyen sevgili annem Emsal Öçal'a, ve babam Şükrü Öçal'a tez sürecimde sağladıkları motivasyon için teşekkür ederim.

EMİNE ÖÇAL

İÇİNDEKİLER

ÇİZELGELER

SİMGELER VE KISALTMALAR

Simgeler

$\pi_{i.}$	: i. satırın marjinal oranı
$\pi_{.j}$	: j. sütunun marjinal oranı
π_{ij}	: Örneklem oranı
y_{ij}	: Gözlenen sıklık
μ_{ij}	: Beklenen sıklık
n	: Örneklem büyüklüğü
α_{ij}	: Y_1 değişkenindeki kayıp durumu için odds
b_{ij}	: Y_2 değişkenindeki kayıp durumu için odds
θ	: Odds oranı
$\hat{\theta}$	: Odds oranının tahmini
χ^2	: Ki-Kare test istatistiği
G^2	: Olabilirlik Oran istatistiği

Kısaltmalar

MAR	: Rastgele kayıp gözlem
MCAR	: Tamamen rastgele kayıp gözlem
NMAR	: Rastgele olmayan kayıp gözlem
EM	: Beklenti maksimizasyonu algoritması
AIC	: Akaike bilgi kriteri
BIC	: Bayesci bilgi kriteri
sd	: Serbestlik derecesi

1. GİRİŞ

1.1. Giriş ve Amaç

Bilimsel çalışmalarda elde edilen veriler genellikle çok değişkenli olup karmaşık yapıya sahiptir. Değişken sayısının fazla olduğu çalışmalarda tüm verilerin eksiksiz bir biçimde toplanması zor olabilmektedir. Bu durum kayıp gözlem ortaya çıkma ihtimalini arttırmaktadır. Kayıp gözlem, veri setinde bir ya da birden fazla hücrede değer olmaması durumudur. Ölçüm cihazının eksik ölçüm yapması, anketin eksik cevaplanması, verilerin yanlışlıkla silinmesi, veri ön işleme sırasında olabilen veri kayıpları gibi durumlar kayıp gözlemlerin ortaya çıkmasına sebep olmaktadır. İstatistiksel yöntemler ve dolayısıyla istatistiksel çözümlemelerde kullanılan yazılımlar verilerin tam olarak elde edildiği varsayımına dayanır. Bu nedenle analizlerde, veri setinde kayıp gözlem olmaması istenir (Allison, 2001).

Kayıp gözlemler, örneklemin rastgelelik özelliğini kaybetmesine ve kestirimlerin yanlı sonuçlar ortaya çıkarmasına neden olabilir. Ayrıca örneklem genişliğinin azalması, dolayısıyla serbestlik derecesinin azalması ve buna bağlı olarak standart hatanın artması da çözümlemelerde güç kaybına neden olabilmektedir (Peng ve ark., 2006; Kale, 2020). Kayıp gözlem içeren verilerde öncelikle bu problemin çözülmesi, daha sonra analizlere geçilmesi daha doğru sonuçlar elde edilmesine olanak sağlar.

Kayıp gözlem sorununu çözebilmek için kayıp durumunun neden kaynaklandığı ve verinin ne kadar kayıp gözlem içerdiği belirlenmelidir (Rubin, 1976). Bu bilgilerden yola çıkarak ilk önce kayıp gözlemin türü belirlenir, daha sonra bu kayıp gözlem türüne yönelik çözüm yolları uygulanabilir.

Kayıp gözlem türleri varlığı ihmal edilebilen ve varlığı ihmal edilemeyen kayıp gözlem olmak üzere iki sınıfa ayrılmıştır (Rubin, 1976; Little ve Rubin, 1987).

Kayıp gözlem problemini çözmek için farklı yöntemler önerilmiştir (Little ve Rubin, 1987). İlk yöntem gözlenen veriye dayalı yöntemlerdir. Bu yöntem kayıp gözlem türünün ihmal edilebilir olduğu durumda veri setinden kayıp gözlemlerin çıkarılması yöntemidir. İkinci yöntem olan yerine koyma yöntemi kayıp gözlem türünün ihmal edilebilir olduğu durumda, verideki değişkenlerin yapısına bağlı olarak önerilen farklı metodlar ile kayıp gözlemlerin yerine değer atanması yöntemidir. Üçüncü yöntem olan modelleme yöntemleri kayıp gözlem türünün ihmal edilemez olduğu durumda uygulanabilmektedir. Bu yönteme göre kayıp gözlemleri modele dahil ederek çıkarsamalar yapılabilmektedir.

Bu çalışmadaki amaç, kayıp gözlem yapısı ve çeşitlerini açıklamak, kayıp gözlemle başa çıkma yöntemlerini tanıtmak ve olumsallık tablolarında kayıp gözlemlerle başa çıkma yöntemlerinden logaritmik doğrusal (log-doğrusal) modelleri incelemektir. Bu çalışma araştırmacıların ve veri analistlerinin kayıp gözlemlerin istatistiksel veri analizi üzerindeki olası etkilerini ve kayıp gözlem problemiyle nasıl başa çıkılabileceğini anlamlarına yardımcı olabilir.

Bu tez çalışması beş bölüme ayrılmaktadır. Birinci bölümde kayıp gözlemlerle ilgili genel bilgiler ve literatürde yer alan çalışmalar hakkında bilgi verilecektir. İkinci bölümde olumsallık tablolarının yapısı ve logaritmik doğrusal modeller hakkında bilgi sağlanacak, kayıp gözlem türleri ve kayıp gözlemlerle başa çıkma yöntemleri tanıtılacaktır. Çalışmanın üçüncü bölümünde kayıp gözlem içeren olumsallık tablolarında logaritmik doğrusal modeller ve kapalı form tahminleri incelenecektir. Dördüncü bölümde gerçek bir veri seti üzerinden uygulama yapılacak ve beşinci bölümde ise sonuçlar tartışılacaktır.

1.2. Önceki Çalışmalar

Kayıp gözlemler, istatistiksel analizlerde problemlerle karşılaşılmasına neden olmaktadır. Bu nedenle, literatürde farklı veri yapıları ve farklı kayıp gözlem çeşitlerinin tartışıldığı birçok çalışma yer almaktadır.

Kayıp gözlemler ile ilgili ilk çalışma Afifi ve Elashof (1966) tarafından yapılan çok değişkenli verilerde kayıp gözlemlerin tartışıldığı çalışmadır. Bu çalışmada çok değişkenli verilerde değişkenlerden birinde ya da birkaçında kayıp gözlemler olması durumunda ortalama, varyans, korelasyon ve doğrusal regresyon parametrelerinin tahmin edilmesi konusu ele alınmıştır.

Kayıp gözlemlerle ilgili en kapsamlı çalışmalar Donald B. Rubin tarafından yapılmıştır. Rubin (1976) çalışmasında kayıp gözlem türlerini sınıflandırmıştır. Hartley (1958) çalışmasında ilk olarak beklenti maksimizasyonu (Expectation Maximization: EM) algoritmasını tanıttıktan sonra Dempster, Laird ve Rubin (1977) çalışmasında kayıp gözlem içeren verilerden en çok olabilirlik tahmini hesaplanmasında kullanılan EM algoritmasını önermiştir. Rubin (1987) çalışmasında kayıp gözlemler için çoklu atama tartışılmıştır. Little ve Rubin (1987) çalışmasında kayıp verilerin olduğu durumda kullanılan istatistiksel yöntemler tartışılmıştır.

Literatürde kayıp gözlem probleminde modelleme yöntemleri konusunda da farklı çalışmalar mevcuttur. Schafer (1997) çalışmasında olumsallık tablolarında kayıp gözlem varlığında parametreleri tahmin etmek için logaritmik doğrusal modeller altında kullanılan farklı algoritmalar önerilmiştir.

Chen ve Frenberg (1976) çalışmasında, kayıp gözlem içeren iki ve çok boyutlu olumsallık tablolarının çözümlenmesinde doygun logaritmik doğrusal modelin parametre tahminlerini hesaplamak amacıyla en çok olabilirlik tahminine dayanan EM algoritması önerilmiştir.

Fuchs (1982) çalışmasında, kayıp gözlem içeren olumsallık tablolarında logaritmik doğrusal modelin en çok olabilirlik tahminini hesaplamak ve en uygun modeli seçmek amacıyla EM algoritması önerilmiştir.

Baker, Rosenberger ve Dersimonian (1992) çalışmasında, kayıp gözlem içeren iki boyutlu olumsallık tablolarında logaritmik doğrusal modellerin kapalı form tahminlerini tanıtılmış ve kapalı form tahminlerinin logaritmik doğrusal modellerin uyumuna etkisi tartışılmıştır.

Phillips (1993) çalışmasında, tamamen rastgele ve rastgele kayıp gözlem varlığında model parametre tahminlerini hesaplamak amacıyla en çok olabilirlik tahmin yöntemi ve EM algoritması ile istatistiksel veri analizi tartışılmıştır.

Perlman ve Wu (1999) çalışmasında, monoton olmayan kayıp gözlemler için koşullu bağımsızlık modeli önerilmiştir. Wu, Reeves ve Perlman (2000) çalışmasında kayıp gözlem içeren olumsallık tabloları için model tahmin edicileri konusu ele alınmış; kayıp gözlemler için ampirik Bayes tahminleri önerilmiştir. Ayrıca bu çalışmada EM algoritmasına alternatif olarak kısıtlı EM algoritması (ER) önerilmiştir.

Molenberghs ve diğerleri (2008) çalışmasında her NMAR modelinin eşit uyumlu bir MAR modeli karşılığı olduğu gösterilmiştir.

Kim, Park ve Kim (2015) çalışmasında, kayıp gözlem içeren çok boyutlu olumsallık tablolarında MAR ve NMAR modelleri tartışılmıştır. Kim, Jeon ve Kim (2020) çalışmasında ise değişkenlerden bir tanesinde MAR ve NMAR türü kayıp gözlem ortaya çıkan iki boyutlu olumsallık tablolarında kullanılabilecek logaritmik doğrusal modeller tartışılmıştır.

Rochani ve diğerleri (2017) çalışmasında, kayıp gözlem içeren üç boyutlu olumsallık tablolarında homojen log-lineer model ve ortak odds oranı tartışılmıştır.

Ghosh ve Vellaisamy (2016, 2019, 2020, 2024) çalışmalarında, kayıp gözlem içeren iki ve üç boyutlu olumsallık tablolarında logaritmik doğrusal modellerin kapalı form tahminleri ve sınır çözümleri tartışılmıştır.

Hooda ve Barak (2019, 2020) çalışmalarında bulanık matrislerde kayıp gözlemleri tahmin etmek için yerine koymaya dayalı bir algoritma önerilmiştir.

Öztürk ve diğerleri (2023) çalışmasında, kayıp gözlem çeşitlerinin ikili sınıflandırmalarda derin öğrenmeye etkisi üzerine bir benzetim çalışması yapılmıştır. Bu çalışmada bootstrap EM algoritması tartışılmıştır.

2. GENEL BİLGİLER

2.1. Olumsallık Tablolarının Yapısı

İki ya da daha fazla kategorik değişkenin çaprazlanması sonucunda olumsallık tabloları elde edilir. Olumsallık tabloları "çapraz tablo" olarak da adlandırılır. İki kategorik değişken arasında ilişki olup olmadığı incelenirken RxC şeklinde gösterilen iki boyutlu olumsallık tabloları; üç kategorik değişken arasında ilişki olup olmadığı incelenirken RxCxK şeklinde gösterilen üç boyutlu (çok boyutlu) olumsallık tabloları kullanılır (Aktaş Altunay ve ark., 2021).

2.1.1. İki Boyutlu Olumsallık Tabloları

Satır değişkeni (Y_1) R düzeye sahip, sütun değişkeni (Y_2) ise C düzeye sahip bir olumsallık tablosu iki boyutlu olarak adlandırılır ve RxC tablosu olarak gösterilir. RxC şeklinde bir olumsallık tablosu Çizelge 2.1.'de gösterilmiştir (Aktaş Altunay ve ark., 2021).

Çizelge 2. 1. RxC olumsallık tablosu

Y_1/Y_2	**1**	**2**	**…**	**C**	**Toplam**
1	y_{11}	y_{12}	$\cdots$	y_{1C}	$y_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2C}	$y_{2.}$
⋮	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
R	y_{R1}	y_{R2}	$\cdots$	y_{RC}	$y_{R.}$
Toplam	$y_{.1}$	$y_{.2}$	$\cdots$	$y_{.C}$	n

İki boyutlu olumsallık tablolarının çözümlenmesinde ihtiyaç duyulan eşitlikler aşağıda verilmiştir.

y_{ij}: i. satır ve j. sütuna karşılık gelen gözlenen sıklık

$\pi_{i.}$: i. marjinal satır oranı

$$\pi_{i.} = \frac{y_{i.}}{n}, \quad i = 1,2,\dots,R \tag{2.1}$$

$\pi_{.j}$: j. marjinal sütun oranı

$$\pi_{.j} = \frac{y_{.j}}{n}, \quad j = 1,2,\dots,C \tag{2.2}$$

π_{ij}: i. satır ve j. sütuna karşılık gelen örneklem oranı

$$\pi_{ij} = \frac{y_{ij}}{n}, \quad i = 1, \dots R;\ j = 1, \dots C \tag{2.3}$$

μ_{ij}: i. satır ve j. sütuna karşılık gelen beklenen sıklık

$$\mu_{ij} = \frac{y_{i.}y_{.j}}{n}, \quad i = 1, \dots R;\ j = 1, \dots C \tag{2.4}$$

İki boyutlu olumsallık tablolarında satır ve sütün değişkenleri arasında ilişki olup olmadığı

H_0: Y_1 ve Y_2 değişkenleri arasında ilişki yoktur.

H_s: Y_1 ve Y_2 değişkenleri arasında ilişki vardır.

şeklinde kurulan hipotez ile test edilir. Bu hipotezin doğruluğu için Eşitlik (2.5)'te yer alan Ki-Kare test istatistiğinden yararlanılır.

$$X^2 = \sum_{i=i}^{R}\sum_{j=2}^{C}\frac{\left(y_{ij} - \mu_{ij}\right)^2}{\mu_{ij}} \tag{2.5}$$

RxC tablosunda serbestlik derecesi (sd):

$$sd = (R - 1)(C - 1)$$

X^2 değeri $X^2_{a;sd}$ değeri ile karşılaştırılır. $X^2 \geq X^2_{a;sd}$ ise hipotez $\%a$ anlamlılık düzeyinde reddedilir. İki değişken arasında anlamlı bir ilişkiden söz edilebilir.

2.1.2. Üç Boyutlu Olumsallık Tabloları

İki boyutlu olumsallık tablolarına bir tabaka değişkeninin eklenmesi ile RxCxK yani üç boyutlu olumsallık tabloları oluşur (Aktaş Altunay ve ark., 2021). Bu tablolarda R ($i = 1,2, \dots, R$) satır değişkeninin düzey sayısını, C ($j = 1,2, \dots, C$) sütun değişkeninin düzey sayısını ve K ($k = 1,2, \dots, K$) tabaka değişkeninin düzey sayısını ifade eder. RxCxK şeklinde bir olumsallık tablolarının çözümlenmesinde kullanılan modeller Çizelge 2.2.'de gösterilmiştir (Bishop, Fienberg ve Holland, 1975).

Çizelge 2. 2. RxCxK olumsallık tablolarında kullanılan modeller

Model	Gösterim	Terim	Serbestlik Derecesi
Bağımsızlık	M0	[1][2][3]	RCK-R-C-K+2
Kısmi Bağımsızlık	M1	[1][23]	RCK-R-CK+1
	M2	[2][13]	RCK-C-RK+1
	M3	[3][12]	RCK-K-RC+1
Koşullu Bağımsızlık	M4	[13][23]	K(R-1)(C-1)
	M5	[12][23]	C(R-1)(K-1)
	M6	[12][13]	R(C-1)(K-1)
Karşılıklı Bağımsızlık	M7	[12][13][23]	(R-1)(C-1)(K-1)
Doygun Model	M8	[12][13][23][123]	0

Çizelge 2.2.'de yer alan modeller için farklı hipotezler test edilmektedir (Aktaş Altunay ve ark., 2021). M0 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{i..}\, \pi_{.j.}\, \pi_{..k}$$

biçiminde kurulur. Satır, sütun ve tabaka değişkeni birbirinden bağımsızdır.

M1 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{i..}\, \pi_{.jk}$$

biçiminde kurulur. Satır değişkeni, sütun ve tabaka değişkeninden bağımsızdır.

M2 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{.j..}\, \pi_{i.k}$$

biçiminde kurulur. Sütun değişkeni, satır ve tabaka değişkeninden bağımsızdır.

M3 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{..k}\, \pi_{ij.}$$

biçiminde kurulur. Tabaka değişkeni, satır ve sütun değişkeninden bağımsızdır.

M4 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{i.k}\, \pi_{.jk} | \pi_{..k}$$

biçiminde kurulur. Tabaka değişkeni sabitken, satır ve sütun değişkeni bağımsızdır.

M5 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{i.j}\, \pi_{.jk} \big| \pi_{.j.}$$

biçiminde kurulur. Sütun değişkeni sabitken, satır ve tabaka değişkeni bağımsızdır.

M6 modeli için hipotez

$$H_0: \pi_{ijk} = \pi_{ij.}\, \pi_{i.k} | \pi_{i..}$$

biçiminde kurulur. Satır değişkeni sabitken, sütun ve tabaka değişkeni bağımsızdır.

M7 modeli için hipotez

$$H_0: \frac{\pi_{111}\pi_{ij1}}{\pi_{i11}\pi_{1j1}} = \frac{\pi_{11k}\pi_{ijk}}{\pi_{i1k}\pi_{1jk}}$$

biçiminde kurulur. İkili etkileşimler birbirinden bağımsızdır.

M8 modeli doygun model olduğu için her durum için $G^2 = 0$ ve $p = 1$ olur, bu nedenle bu model için hipotez testi yapılmaz.

Tüm modeller için genel hipotez olan

H_0: Modele uyum vardır

hipotezini test edebilmek için her bir model varsayımı altında beklenen sıklıklar (μ_{ijk}) hesaplanır. Modele uyumu araştırmak amacıyla beklenen sıklıklardan yararlanılarak hesaplanan Ki-Kare (χ^2) ve Olabilirlik Oran (G^2) test istatistikleri kullanılır. Test istatistikleri Eşitlik 2.6 ve Eşitlik 2.7'de özetlenmiştir (Aktaş Altunay ve ark., 2021).

$$X^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{K}\frac{\left(y_{ijk} - \mu_{ijk}\right)^2}{\mu_{ijk}} \tag{2.6}$$

$$G^2 = 2\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{K} y_{ijk}\, ln\left(\frac{y_{ijk}}{\mu_{ijk}}\right) \tag{2.7}$$

Hipotez testleri sonucunda uyum bulunan modellerden en iyi modeli seçmek için Eşitlik 2.8'de yer alan Akaike (AIC) ve Eşitlik 2.9'da yer alan Bayesci (BIC) bilgi kriterleri hesaplanır.

$$AIC = G^2 - 2\,sd \tag{2.8}$$

$$BIC = G^2 - sd\,ln(n) \tag{2.9}$$

AIC ya da BIC değeri en küçük olan model en iyi modeldir. Parametre tahminleri en iyi modele göre yapılır (Aktaş Altunay ve ark., 2021).

2.2. Logaritmik Doğrusal Modeller

Birden çok kategorik değişkenin çaprazlanması sonucunda olumsallık tabloları elde edilir. Bu tablolarda yer alan değişkenler arasındaki ilişkiler logaritmik doğrusal modeller yardımıyla incelenir (Agresti, 2002).

İki boyutlu olumsallık tablolarında çözümleme yaparken Pearson'ın Ki-Kare istatistiği ve Olabilirlik Oran testi uygulanır. Üç ya da daha çok boyutlu olumsallık tablolarından iki değişkenli marjinal tablolar oluşturarak her birine ki-kare testi uygulayarak çözümleme yapmak büyük oranda bilgi kaybına neden olabilir. Böyle durumlarda logaritmik doğrusal modeller yardımıyla olumsallık tablolarındaki göze sıklıkları modellenerek kategorik değişkenler arasındaki ilişkiyi incelemek mümkündür (Heijden, Falguerolles ve Leeuw, 1989; Çağılcı ve Danacıoğlu, 2020). RxCxK şeklinde gösterilen üç boyutlu tablolarda logaritmik doğrusal model eşitlikleri Çizelge 2.3.'te özetlenmiştir (Aktaş Altunay ve ark, 2021).

Çizelge 2. 3. RxCxK tabloları için logaritmik doğrusal modeller

Model	**Model Denklemi**
1,2,3	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3$
1,23	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{jk}^{23}$
2,13	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ik}^{13}$
3,12	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12}$
13,23	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ik}^{13} + \lambda_{jk}^{23}$
12,23	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{jk}^{23}$
12,13	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13}$
12,13,23	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23}$
123	$log\mu_{ijk} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123}$

Veri setinde kayıp gözlemler söz konusu olduğunda logaritmik doğrusal modeller doğrudan uygulanamaz. Analizlere geçilmeden önce kayıp gözlem probleminin çözülmesi gerekir.

2.3. Kayıp Gözlem

Kayıp gözlem, veri setinde bir ya da daha fazla değişkene ait hücrelerde gözlem değeri bulunmaması durumudur. Kayıp gözlem durumu farklı nedenlerden ortaya çıkabilir, örneğin; yanıtlayıcının anketteki bazı sorulara cevap vermemesi, ölçüm cihazının bozuk olması, veri girişinde yapılan hatalar kayıp gözlem sorununa neden olabilmektedir. İstatistiksel analizlerde güvenilir sonuçlar elde edebilmek için analizlerden önce kayıp gözlem incelemesi yapılmalı ve uygun yöntemlerle kayıp gözlem sorunu çözülmelidir. Bu bölümde kayıp gözlem türleri ve uygun çözüm yöntemleri ele alınacaktır.

Kayıp gözlem sorununu çözebilmek için kullanılacak yöntem kayıp gözlemin yapısına ve türüne bağlı olarak değişiklik gösterir. Rubin (1976) çalışmasında kayıp gözlemlerin ortaya çıkma şekline göre bir sınıflandırma önerilmiştir. Bu sınıflandırma daha sonra yapılan tüm çalışmalar için bir öncü olmuştur. Bu tanımlamaya göre rastgele kayıp gözlem, tamamen rastgele kayıp gözlem ve rastgele olmayan kayıp gözlem olmak üzere üç çeşit kayıp gözlem vardır. Rastgele kayıp gözlem ve tamamen rastgele kayıp gözlem “varlığı ihmal edilebilen kayıp gözlem” olarak nitelendirilirken; rastgele olmayan kayıp gözlem ise “varlığı ihmal edilemeyen kayıp gözlem” olarak nitelendirilmektedir (Little ve Rubin, 1987).

2.3.1.Varlığı İhmal Edilebilen Kayıp Gözlem

İhmal edilebilen kayıp gözlem, bir veri setindeki kayıp gözlemlerin analiz veya sonuçlar üzerindeki etkisinin istatistiksel olarak anlamlı düzeyde olmaması durumudur. Bu durumda genellikle kayıp gözlem sayısı, örneklem büyüklüğüne göre oldukça düşüktür.

Varlığı ihmal edilebilen kayıp gözlem türleri ikiye ayrılmaktadır.

1. Rastgele kayıp gözlem (MAR: Missing at random)
2. Tamamen rastgele kayıp gözlem (MCAR: Missing completely at random)

2.3.1.1. Rastgele Kayıp Gözlem

X ve Y değişkenlerinden Y değişkeninde kayıp gözlemler mevcutsa ve cevap olasılıkları X değişkenine bağlıysa rastgele kayıp gözlem durumu ortaya çıkar (Little ve Rubin, 1987; Pembegül, 2009).

Rastgele kayıp gözlem (MAR), verideki kayıp gözlemlerin kendi değişkenine değil, gözlenen diğer değişkenlere bağlı olduğu durumdur. Dolayısıyla bu durumda yalnızca gözlenen sıklıklar üzerinden çıkarsama yapılabilir (Little ve Rubin, 1987). Örneğin, bir araştırmacı bir bölümdeki öğrencilerin A dersinden aldıkları sınav notlarını öğrenmek istesin. Öğrencilerin bir kısmı A dersinin sınavına girmemiş ya da araştırmacıya sınav sonucunu bildirmemiş olsun. Bu durum A dersine düzenli katılma, derste düzenli not tutma gibi diğer değişkenlerle ilişkilendirilebilir. Derse düzenli katılamayan ve ders notları eksik olan öğrenciler yeterli çalışamamış ve sınava girmemiş olabilir ya da düşük not aldıkları için araştırmacıyla sınav sonucunu paylaşmak istememiş olabilir. MAR durumundan söz edebilmek için ilişkilendirilecek diğer değişkenlerin mevcut olması ve bu değişkenlerin tamamen gözlenebilir olması gerekmektedir.

MAR durumu Eşitlik (2.10)'da yer almaktadır (Howell, 2007).

$$P(Y\ kayıp|Y,X) = P(Y\ kayıp|X) \tag{2.10}$$

2.3.1.2. Tamamen Rastgele Kayıp Gözlem

X ve Y değişkenlerinden Y değişkeninde kayıp gözlemler mevcutsa ve cevap olasılıkları X ve Y değişkenlerinden bağımsızsa tamamen rastgele kayıp gözlem durumu ortaya çıkar (Little ve Rubin, 1987; Pembegül, 2009).

Tamamen rastgele kayıp gözlem (MCAR), verideki kayıp gözlemlerin hiçbir değişkene bağlı olmadan tamamen rastgele bir şekilde kayıp olması durumudur (Little ve Rubin, 1987). MCAR türü kayıp gözlemler MAR'nin özel bir durumudur. Bir veri setinde kayıp gözlem türü MCAR ise aynı zamanda MAR'dr denebilir (Allison, 2001). Katılımcının anketteki bir soruyu görmeyip boş bırakması, anketin kaybolması, laboratuvardaki bir örneğin hasar görmesi gibi durumlar MCAR türü kayıp gözleme örnektir.

MCAR türü dışındaki kayıp gözlemlerin sonuçlarının yanlı olma olasılığı fazladır. Bu nedenle MCAR türü dışındaki kayıp gözlemlerin çözümünde daha sağlam yöntemler kullanılmalıdır (Little ve Rubin, 1987; Yazıcı, 2005; Köse ve Öztemur, 2014; Alpar, 2021).

MCAR durumu Eşitlik (2.11)'de yer almaktadır (Allison, 2001).

$$P(Y\ kayıp|Y,X) = P(Y\ kayıp) \tag{2.11}$$

2.3.2. Varlığı İhmal Edilemeyen (Rastgele Olmayan) Kayıp Gözlem

X ve Y değişkenlerinden Y değişkeninde kayıp gözlemler mevcutsa ve cevap olasılıkları Y değişkeninin kendisine bağlıysa rastgele olmayan kayıp gözlem durumu ortaya çıkar (Little ve Rubin, 1987; Pembegül, 2009).

Bir veri setinde bulunan kayıp gözlemler rastgele değil de sistemli olarak belirli bir düzen içinde ortaya çıkıyorsa varlığı ihmal edilemeyen kayıp gözlemdir. Bu durum genellikle araştırma sürecindeki tasarım, uygulama ya da ölçüm hatalarından kaynaklanmaktadır. Varlığı ihmal edilemeyen kayıp gözlem “rastgele olmayan kayıp gözlem (NMAR: Not missing at random) olarak tanımlanmaktadır.

NMAR türü kayıp gözlemde, kayıplar değişkenin kendisiyle ya da veri setinde yer almayan başka değişkenlerle ilgilidir. Örneğin, bir ankette belirli bir grup insanın bir soruyu yanıtlamaktan özellikle kaçınması veya belirli bir özellik için ölçüm yapılmamış olması gibi durumlarda NMAR türü kayıp gözlemden bahsedilebilir.

Çalışmalarda NMAR türü kayıp gözlemler göz ardı edilemez. Bu durumda, kayıp gözlem problemiyle baş edebilmek için kayıp gözlemleri temsil eden uygun bir model kurulması gerekmektedir ve bu da çoğu zaman zorlayıcı olabilmektedir (Howell, 2007). Freedman ve Wachter (2007) çalışmasında nüfus sayımına ilişkin verilerdeki kayıp gözlemler incelemiş ve kayıp gözlemlerde modelleme yapmanın çoğu durumda oldukça zor, bazı durumlarda ise başarısız olabileceğini belirtilmiştir. Çizelge 2.4’te farklı kayıp gözlem türlerinin karşılaştırılması verilmiştir (Dayanıklı, 2021).

Çizelge 2. 4. Farklı kayıp gözlem türlerinin karşılaştırılması

Kayıp Gözlem Türü	**Bağlantısı**	**İhmal Edilebilirlik**	**Kayıp Miktarı**
MCAR	Hiçbir değişkene bağlı değil	Edilebilir	Genelde az
MAR	Gözlenen başka bir değişkene bağlı	Edilebilir	Genelde çok
NMAR	Değişkenin kendisine bağlı	Edilemez	Genelde çok

2.4. Kayıp Gözlem Sorununun Çözümüne Yönelik Yaklaşımlar

Kayıp gözlem problemi analizlerde hatalı çıkarsamalara neden olabildiği için veri analizinden önce bu problemi ortadan kaldırmak gerekir. Literatürde bu konuyla ilgili farklı yöntemler sunulmuştur. Bu yöntemler değişkenin yapısı, kayıp gözlemlerin türü ve sayısı gibi faktörlere bağlı olarak değişmektedir. Verinin kayıp olma nedeni, çözüm için hangi yöntemin seçileceğinin belirlenmesinde önemli bir role sahiptir (Howell, 2007).

Little ve Rubin (1987) çalışmasında kayıp gözlem problemini çözmek için geliştirilen yöntemleri aşağıdaki ana başlıklar altında incelemiştir:

1. Gözlenen veriye dayalı yöntemler (silme yöntemleri)
2. Yerine koyma (değer atama) yöntemleri
3. Modelleme yöntemleri

2.4.1. Gözlenen Veriye Dayalı Yöntemler (Silme Yöntemleri)

Veri analizlerinde kayıp gözlemlere yönelik en yaygın yaklaşım silme yöntemleridir. Bu yöntemler "gözlenen veriye dayalı yöntemler" olarak adlandırılmaktadır. İstatistiksel yazılımların büyük bir kısmı kayıp gözlem durumlarında bu varsayımla çalışmaktadır.

Silme yöntemlerinin uygulanabilmesi için verideki kayıp gözlem türünün varlığı ihmal edilebilen (rastgele ya da tamamen rastgele) kayıp gözlem olması gerekmektedir. Silme yöntemleri, rastgele olmayan kayıp gözlemler için uygun değildir (Howell, 2007).

Silme yöntemleri vaka bazında silme ve çift yönlü silme olarak uygulanmaktadır (Allison, 2001; Howell, 2007).

Vaka bazında silme yöntemine göre kayıp gözlemlerin mevcut olduğu vakalarda (deneklerde), o vakaya (deneğe) ait gözlenenler de dahil olmak üzere tüm veriler silinir. Böylece veride kayıp gözlem tamamen ortadan kalkar. Örneğin, beş değişkenin yer aldığı bir ankette bir katılımcının 1., 3. ve 4. değişkenlere ait yanıtları mevcut fakat 2. ve 5. değişkene ait yanıtları kayıpsa bu katılımcının tüm değerleri silinir. Bu çalışmada yalnızca tüm sorulara yanıt veren katılımcılar üzerinden veri analizi yapılır. Vaka bazında silme yöntemi uygulama kolaylığı bakımından avantajlı olsa da örneklem sayısı azalacağından dolayı, standart hata artar ve testin gücü azalır. Öte yandan önemli miktarda veri kaybı yaşanabilir ve bu da tahminlerin yanlı olmasına neden olabilir.

Çift yönlü silme yönteminde yalnızca gözlenen değerlerin bulunduğu değişkenler kullanılır. Örneğin, beş değişkenin yer aldığı bir ankette bir katılımcının 1., 3. ve 4. değişkenlere ait yanıtları mevcut fakat 2. ve 5. değişkene ait yanıtları kayıpsa bu katılımcının gözlenen 1.,3. ve 4. değişkene ait değerleri veri analizine dahil edilir.

Çift yönlü silme yönteminde vaka bazında silme yöntemine göre veri kaybı daha az olur. Bu nedenle testin gücü de daha yüksektir. Ancak otalama, standart sapma ve standart hata değerlerini doğru tahmin etmek zordur. Ayrıca değişkenler arasında ilişki miktarı yüksek ise bu yöntem yanlı sonuçlara neden olabilir. Öte yandan, bazı durumlarda varyans-kovaryans matrislerinin negatif çıkması nedeniyle terslerinin hesaplanması mümkün olmayabilir ve gerekli parametreler hesaplanamayabilir (Allison, 2001; Howell, 2007).

2.4.2. Yerine Koyma (Değer Atama) Yöntemleri

Yerine koyma yöntemi, kayıp gözlemlerin yerine basit yöntemler ya da gelişmiş teknikler kullanılarak uygun değerlerin atanması işlemidir. Örneğin, kayıp gözlemlerin bulunduğu hücrelere ortalama, tepe değeri, ortanca ya da sabit bir değer ataması yapılabilir. Bu yöntem "imputasyon metodu" olarak da adlandırılmaktadır.

Değer atama yöntemleri de genellikle silme yöntemleri gibi varlığı ihmal edilebilen kayıp gözlem durumlarında uygulanabilmektedir. Kayıp gözlemin bulunduğu değişkenin yapısına ve kayıp gözlem sayısına göre aşağıda yer alan atama yöntemleri kullanılmaktadır (Rubin, 1987):

- Hot-Deck yerine koyma yöntemi
- Cold-Deck yerine koyma yöntemi
- Değiştirme yerine koyma yöntemi
- Ortalama yerine koyma yöntemi
- Regresyon yerine koyma yöntemi
- Stokastik yerine koyma yöntemi
- Birleşik yerine koyma yöntemi
- Çoklu yerine koyma yöntemi

2.4.3. Modelleme Yöntemleri

Silme ve yerine koyma yöntemleri genel olarak varlığı ihmal edilebilen kayıp gözlem türleri için çözüm sağlamaktadır. Varlığı ihmal edilemeyen kayıp gözlem türlerinde ise modelleme yöntemlerinin kullanılması önerilmiştir (Little ve Rubin, 1987; Pembegül, 2009). Modelleme yöntemleri kayıp gözlemlerin etkisinin en aza indirilmesi için kalıcı çözümler sağlayabildiği için diğer yöntemlere göre oldukça avantajlıdır Little ve Rubin (1987) çalışmasında modelleme yöntemlerini aşağıdaki gibi sınıflandırmıştır:

I. Stokastik modelleme yöntemleri

II. Bayesci modelleme yöntemleri

III. Logaritmik doğrusal modeller

Stokastik modelleme yöntemleri genellikle boylamsal verilerde aralıklı olarak kayıp gözlemlerin bulunması durumunda kullanılmaktadır. Bu tür kayıp gözlemlerin varlığında, modele ait parametre tahminlerini bulabilmek için Stokastik EM (SEM) algoritması (Celuex ve Diebolt, 1985) ve parametre tahminlerinin standart hatalarını bulabilmek için de Monte Carlo yöntemi gibi yaklaşımlar uygulanmaktadır (Gad ve Ahmed, 2006).

Bayesci modelleme yöntemlerinde ise kayıp gözlem içeren olumsallık tablolarında değişkenlere ait önsel bilgiler kullanılarak parametre çıkarsamaları yapılabilmektedir. Bayesci modelleme yaklaşımı için literatürde farklı çalışmalar mevcuttur. Bunlardan biri kayıp gözlemlerin ve gözlenen değerlerin aynı model altında incelendiği ampirik Bayes modelleme yaklaşımıdır (Park ve Brown, 1994; Park, 1998). Bu yaklaşım küçük tablolar için kullanışlı olsa da çok boyutlu tablolar için bazı durumlarda uygun bulunmamıştır (Park ve Choi, 2010). Çok boyutlu olumsallık tablolarında Monte Carlo Markov Chain (MCMC) yöntemi ve modellerde sonsal dağılımları maksimize etmek için EM algoritması önerilmiştir (Park ve Choi, 2010).

Bu tez kapsamında kayıp gözlem varlığında kullanılan logaritmik doğrusal modeller incelenmiştir. Logaritmik doğrusal modellere ilişkin kapsamlı bilgiler tezin üçüncü bölümünde sunulmuştur.

3. LOGARİTMİK DOĞRUSAL MODELLER

Kayıp gözlem içeren olumsallık tablolarında modelleme yöntemlerini kullanarak farklı modellere ait parametreleri tahmin etmek mümkündür (Rochani ve diğerleri, 2015). Modelleme yöntemlerine yönelik en kapsamlı çalışmalar en çok olabilirlik tahminlerine dayalı logaritmik doğrusal modeller ile yapılmıştır (Pembegül, 2009). Bu yöntemlerde amaç verileri iyi bir şekilde temsil eden modelin bulunmasıdır.

3.1. Kayıp Gözlem İçeren İki Boyutlu Olumsallık Tabloları İçin Logaritmik Doğrusal Modeller

Y_1 ve Y_2 sırasıyla R ve C düzeylerine sahip iki kategorik değişken olmak üzere yalnızca tek değişkende (satır veya sütun değişkeninde) kayıp olması ve iki değişkende (hem satır hem sütun değişkenlerinde) kayıp olması durumları incelenmiştir (Baker, Rosenberger ve Dersimoian, 1992).

3.1.1. Tek Değişkende Kayıp Olması Durumu

Y_2 (sütun) değişkeninde kayıp gözlemler mevcut olsun. Kayıp durumu, Y_2 gözleniyorsa $M = 1$, Y_2 kayıpsa $M = 2$ olmak üzere M adında yeni bir değişken ile gösterilsin. Bu durumda olumsallık tablosu Çizelge 3.1'deki gibi oluşturulur (Kim, Park ve Kim, 2015).

Çizelge 3. 1. Tek değişkende kayıp olan iki boyutlu olumsallık tablosu

	M = 1				**M = 2**
	Y_2 = 1	**Y_2 = 2**	⋯	**Y_2 = C**	**Kayıp**
Y_1 = 1	y_{111}	y_{121}	⋯	y_{1C1}	y_{1+2}
Y_1 = 2	y_{211}	y_{221}	⋯	y_{2C1}	y_{2+2}
⋮	⋮	⋮	⋮	⋮	⋮
Y_1 = R	y_{R11}	y_{R21}	⋯	y_{RC1}	y_{R+2}

Çizelge 3.1'de yer alan olumsallık tablosuna ait eşitlikler aşağıdaki gibi tanımlanır (Kim, Park ve Kim, 2015; Ghosh ve Vellaisamy, 2024):

$i = 1,2, \dots, R,\ j = 1,2, \dots, C\ \ ve\ \ x = 1, 2$ olmak üzere göze sıklıkları $\{y_{ijx}\}$ şeklinde gösterilmiştir.

Gözlenen sıklıklar vektörü $y_{göz} = (\{y_{ij1}\}, \{y_{i+2}\})$ olmak üzere $\{y_{ij1}\}$ tamamen gözlenen; $\{y_{i+2}\}$ ise Y_2 değişkeninin kayıp olduğu durumdaki değerlerden oluşmaktadır. Burada gösterilen "+" işareti Y_2 değişkeni kayıp olduğu için tüm durumlardaki toplamı ifade etmektedir.

Örneklem oranları vektörü $\pi = \{\pi_{ijx}\}$ ve beklenen sıklıklar vektörü $\mu = \{\mu_{ijx}\}$ şeklinde gösterilir. $n = \sum_{i,j,x} y_{ijx}$ olmak üzere toplam örneklem büyüklüğünü ifade etmektedir.

Çizelge 3.1'de verilen olumsallık tablosu için $i = 1,2,\dots,R,\ j = 1,2,\dots,C$ ve $x = 1,2$ olmak üzere logaritmik doğrusal model Eşitlik (3.1)'deki gibi tanımlanır.

$$log\mu_{ijx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M} + \lambda_{jx}^{Y_2M} \tag{3.1}$$

Bu modeldeki tüm parametreler, parametre tahminleri toplamının sıfır olması koşulunu sağlar. Örneğin, $\sum_i \lambda_{Y_1Y_2(i,j)} = \sum_j \lambda_{Y_1Y_2(i,j)} = 0$

Y_1 değişkenindeki kayıp türünü belirlemek için odds Eşitlik (3.2)'de verilmiştir.

$$b_{ij} = \frac{P(M = 2 \mid Y_1 = i, Y_2 = j)}{P(M = 1 \mid Y_1 = i, Y_2 = j)} = \frac{\pi_{ij2}}{\pi_{ij1}} = \frac{\mu_{ij2}}{\mu_{ij1}} \tag{3.2}$$

Beklenen sıklıklar Eşitlik (3.3)'te verilmiştir.

$$\mu_{ij2} = b_{ij}\mu_{ij1} \tag{3.3}$$

Beklenen sıklar Eşitlik (3.4)'te yer alan koşulu sağlar.

$$\sum_{ij} \mu_{ij1}(1 + b_{ij}) = n \tag{3.4}$$

Birleşik olasılıklar Eşitlik (3.5)'teki gibi hesaplanır ve buradan marjinal olasılıklar elde edilebilir.

$$\pi_{ij+} = \frac{\mu_{ij1}(1 + b_{ij})}{n} \tag{3.5}$$

Eşitlik (3.1)'den yararlanılarak

$$b_{ij} = exp[-2\{\lambda_M(1) + \lambda_{Y_1M}(i,1) + \lambda_{Y_2M}(j,1)\}] \tag{3.6}$$

elde edilebilir.

b_{ij}, kayıp gözlem türlerine göre aşağıdaki üç durumda incelenebilir (Baker ve Laird, 1988; Kim, Park ve Kim, 2015).

1) $b_{ij} = b_{.j}$ ise NMAR türü kayıp gözlem (kayıp durumu yalnızca Y_2'ye yani değişkenin kendisine bağlıdır.)
2) $b_{ij} = b_{i.}$ ise MAR türü kayıp gözlem (kayıp durumu Y_1 değişkenine bağlıdır.)
3) $b_{ij} = b_{..}$ ise MCAR türü kayıp gözlem (kayıp durumu değişkenlerden bağımsızdır.)

Δ bir sabit olmak üzere, Eşitlik (3.7)'de Poisson örneklemesi altında μ'nün log-olabilirlik fonksiyonu verilmiştir.

$$l(\mu; y_{göz}) = \sum_{i,j} y_{ij1} log\mu_{ij1} + \sum_{i} y_{i+2} log\mu_{i+2} - \sum_{i,j,x} \mu_{ijx} + \Delta \qquad (3.7)$$

Kayıp gözlemlerin varlığında kullanılan logaritmik doğrusal modeller serbestlik derecesinden daha fazla parametre içerdiği için tanımlanabilir değildir (Baker, Rosenberger ve Dersimonian, 1992). Bu nedenle modellerin kapalı formlarının kullanılarak yeniden parametrelendirmesi önerilmiştir.

Yukarıda tanımlanan NMAR, MAR ve MCAR türü kayıp gözlemlerin olduğu üç durum için logaritmik doğrusal modeller ve bu modellerin kapalı form tahminleri aşağıda verilmiştir:

Model 1: $b_{ij} = b_{.j}$ (Y_2 NMAR)

Y_2 değişkenindeki kayıp gözlemlerin yalnızca Y_2'ye yani değişkenin kendisine bağlı olduğu durumdur. $i = 1,2,\dots,R, j = 1,2,\dots,C$ ve $x = 1, 2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.8)'de yer almaktadır.

$$log\mu_{ijx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M} \qquad (3.8)$$

$i \geq 1$ olmak üzere, $\hat{\beta}_{.j}$ parametresi $\sum_j \hat{\mu}_{ij1}\hat{\beta}_{.j} = y_{i+2}$ koşulunu sağlar.

Beklenen sıklıklar Eşitlik (3.9)'da yer almaktadır.

$$\hat{\mu}_{ij1} = y_{ij1} \qquad (3.9)$$

Model 2: $\boldsymbol{b_{ij} = b_{i.}}$ (Y2 MAR)

Y_2 değişkenindeki kayıp gözlemlerin Y_1 değişkenine bağlı olduğu durumdur. $i = 1,2, \ldots, R$, $j = 1,2, \ldots, C$ ve $x = 1, 2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.10)'da yer almaktadır.

$$log\mu_{ijx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M} \qquad (3.10)$$

Model parametresi Eşitlik (3.11) yardımıyla hesaplanabilir.

$$\hat{\beta}_{i.} = \frac{y_{i+2}}{y_{i+1}}, \qquad i = 1,2, \ldots, R \qquad (3.11)$$

Beklenen sıklıklar Eşitlik (3.12)'de yer almaktadır.

$$\hat{\mu}_{ij1} = \frac{y_{ij1}y_{i++}y_{i+1}}{y_{i+1}y_{i++}} = y_{ij1} \qquad (3.12)$$

Model 3: $\boldsymbol{b_{ij} = b_{..}}$ (Y2 MCAR)

Y_2 değişkenindeki kayıp gözlemlerin iki değişkenlerden de bağımsız olduğu durumdur. $i = 1,2, \ldots, R,\ j = 1,2, \ldots, C$ ve $x = 1, 2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.13)'te yer almaktadır.

$$log\mu_{ijx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} \qquad (3.13)$$

Model parametresi Eşitlik (3.14) yardımıyla hesaplanabilir.

$$\hat{\beta}_{..} = \frac{y_{++2}}{y_{++1}} \qquad (3.14)$$

Beklenen sıklıklar Eşitlik (3.15)'te yer almaktadır.

$$\hat{\mu}_{ij1} = \frac{y_{ij1}y_{i++}y_{++1}}{y_{i+1}y_{+++}} \qquad (3.15)$$

Yukarıda verilen üç model için hipotezler:

H_0: Modele uyum vardır.

H_s: Modele uyum yoktur.

şeklinde kurulur.

Bu hipotezleri test edebilmek için olabilirlik oran istatistiği Eşitlik (3.16)'da verilmiştir.

$$G^2 = -2\left[\sum_{i,j} y_{ij1} ln\left(\frac{\hat{\mu}_{ij1}}{y_{ij1}}\right) + \sum_{i} y_{i+2} ln\left(\frac{\sum_j \hat{\mu}_{ij1}\hat{\beta}_{ij}}{y_{i+2}}\right) - \sum_{i,j} \hat{\mu}_{ij1}\left(1 + \hat{\beta}_{ij}\right) + n\right] \quad (3.16)$$

Bu üç model için serbestlik dereceleri:

$sd = R(C+1)$ – modelde tahmin edilen parametre sayıs

olarak hesaplanır. Modellerde tahmin edilen parametre sayıları Çizelge 3.2.'de verilmiştir.

Çizelge 3. 2. İki boyutlu tablolarda tek değişkende kayıp durumunda modellerden tahmin edilen parametre sayıları

Model	**Parametre Sayısı**
$(b_{i.})$	$RC + R$
$(b_{.j})$	$RC + C$
$(b_{..})$	$RC + 1$

G^2 değeri $X^2_{a;sd}$ değeri ile karşılaştırılır. $G^2 < X^2_{a;sd}$ ise hipotez $\%a$ anlamlılık düzeyinde reddedilemez. Modele uyum vardır.

3.1.2. İki Değişkende Kayıp Olması Durumu

Hem Y_1 (satır) ve hem de Y_2 (sütun) değişkenlerinde kayıp gözlemler mevcut olsun. Y_1 değişkenindeki kayıp durumu, Y_1 gözleniyorsa $M_1 = 1$, Y_2 kayıpsa $M_1 = 2$ olmak üzere M_1 adında yeni bir değişken ile; Y_2 değişkenindeki kayıp durumu ise, Y_2 gözleniyorsa $M_2 = 1$, Y_2 kayıpsa $M_2 = 2$ olmak üzere M_2 adında yeni bir değişken ile gösterilsin. Bu durumda olumsallık tablosu Çizelge 3.3'teki gibi olur (Baker, Rosenberger ve Dersimonian, 1992).

Çizelge 3. 3. İki değişkende kayıp olan iki boyutlu olumsallık tablosu

		M_2 = 1				**M_2 = 2**
		Y_2 = 1	**Y_2 = 2**	⋯	**Y_2 = C**	**Kayıp**
M_1 = 1	**Y_1 = 1**	y_{111}	y_{121}	⋯	y_{1C1}	y_{1+2}
	Y_1 = 2	y_{211}	y_{221}	⋯	y_{2C1}	y_{2+2}
	⋮	⋮	⋮	⋮	⋮	⋮
	Y_1 = R	y_{R11}	y_{R21}	⋯	y_{RC1}	y_{R+2}
M_1 = 2	**Kayıp**	y_{+121}	y_{+221}	⋯	y_{+C21}	y_{++22}

Çizelge 3.3'te verilen olumsallık tablosuna ait eşitlikler aşağıdaki gibi tanımlanır (Baker, Rosenberger ve Dersimonian, 1992).

$i = 1,2, \dots, R,\ j = 1,2, \dots, C$ ve $x, s\ =\ 1, 2$ olmak üzere göze sıklıkları $\{y_{ijxs}\}$ şeklinde gösterilmiştir.

Gözlenen sıklıklar vektörü $y_{göz} = (\{y_{ij11}\}, \{y_{i+12}\}, \{y_{+j21}\}, y_{++22})$ olmak üzere $\{y_{ij11}\}$ tamamen gözlenen sıklıkları göstermektedir. $\{y_{+j21}\}$ Y_1 değişkeninin kayıp olduğu durumdaki; $\{y_{i+12}\}$ Y_2 değişkeninin kayıp olduğu durumdaki ve $\{y_{++22}\}$ ise hem Y_1 hem de Y_2 değişkenlerinin kayıp olduğu durumdaki değerlerden oluşmaktadır. Burada gösterilen "+" işareti kayıp değişkenlerde tüm durumlardaki toplamı ifade etmektedir.

Örneklem oranları vektörü $\pi = \{\pi_{ijxs}\}$ ve beklenen sıklıklar vektörü $\mu = \{\mu_{ijxs}\}$ şeklinde gösterilir. n ise örneklem büyüklüğüdür. $i = 1,2, \dots, R, j = 1,2, \dots, C$ ve $x, s\ =\ 1, 2$ olmak üzere bu olumsallık tablosu için logaritmik doğrusal model Eşitlik (3.17)'deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{jx}^{Y_2M_1} + \lambda_{is}^{Y_1M_2} + \lambda_{js}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \tag{3.17}$$

Bu modeldeki tüm parametreler, parametre tahminleri toplamının sıfır olması koşulunu sağlar. Örneğin, $\sum_i \lambda_{Y_1Y_2(i,j)} = \sum_j \lambda_{Y_1Y_2(i,j)} = 0$

Y_1 ve Y_2 değişkenlerindeki kayıp gözlemler için oddslar sırasıyla α_{ij} ve b_{ij} şeklinde gösterilir.

Y_2 gözlendiğinde, Y_1 değişkeninde kayıp olması durumu için odds Eşitlik (3.18) yardımıyla hesaplanabilir.

$$\alpha_{ij} = \frac{P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j)}{P(M_1 = 1, M_2 = 1 \mid Y_1 = i, Y_2 = j)} = \frac{\pi_{ij21}}{\pi_{ij11}} = \frac{\mu_{ij21}}{\mu_{ij11}} \tag{3.18}$$

Y_1 gözlendiğinde, Y_2 değişkeninde kayıp olması durumu için odds Eşitlik (3.19) yardımıyla hesaplanabilir.

$$b_{ij} = \frac{P(M_1 = 1, M_2 = 2 \mid Y_1 = i, Y_2 = j)}{P(M_1 = 1, M_2 = 1 \mid Y_1 = i, Y_2 = j)} = \frac{\pi_{ij12}}{\pi_{ij11}} = \frac{\mu_{ij12}}{\mu_{ij11}} \tag{3.19}$$

Beklenen sıklıklar $\mu_{ij11} = n\pi_{ij11}$ eşitliğini sağlar.

M_1 ve M_2 arasındaki odds oranı (θ) Eşitlik (3.20) yardımıyla hesaplanabilir.

$$\theta = \frac{P(M_1 = 1, M_2 = 1 \mid Y_1 = i, Y_2 = j)P(M_1 = 2, M_2 = 2 \mid Y_1 = i, Y_2 = j)}{P(M_1 = 1, M_2 = 2 \mid Y_1 = i, Y_2 = j)P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j)}$$

$$= \frac{\pi_{ij11}\pi_{ij22}}{\pi_{ij12}\pi_{ij21}} = \frac{\mu_{ij11}\mu_{ij22}}{\mu_{ij12}\mu_{ij21}} \tag{3.20}$$

Eğer $\theta = 1$ ise M_1 ve M_2 bağımsızdır.

Beklenen sıklıklar Eşitlik (3.21) – Eşitlik (3.23) yardımıyla hesaplanabilir.

$$\mu_{ij21} = \alpha_{ij}\mu_{ij11} \tag{3.21}$$

$$\mu_{ij12} = b_{ij}\mu_{ij11} \tag{3.22}$$

$$\mu_{ij22} = \mu_{ij11}\alpha_{ij}b_{ij}\theta \tag{3.23}$$

Toplam örneklem büyüklüğü Eşitlik (3.24) kullanılarak hesaplanabilir.

$$n = \sum_{i,j} \mu_{ij11}\langle 1 + \alpha_{ij} + b_{ij} + \alpha_{ij}b_{ij}\theta \rangle \tag{3.24}$$

Birleşik olasılık Eşitlik (3.25)'teki gibi gösterilir ve buradan marjinal olasılıklar elde edilebilir.

$$\pi_{ij++} = \frac{\mu_{ij11}\left(1 + \alpha_{ij} + b_{ij} + \alpha_{ij}b_{ij}\theta\right)}{n} \tag{3.25}$$

Y_2 gözlenirken Y_1 değişkeninde kayıp olması olasılığı Eşitlik (3.26) ile, Y_1 gözlenirken Y_2 değişkeninde kayıp olması olasılığı Eşitlik (3.27) ile hesaplanabilir.

$$\phi_{1|2}(i,j) = P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j) = \frac{\alpha_{ij}}{1 + \alpha_{ij}} \tag{3.26}$$

$$\phi_{2|1}(i,j) = P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j) = \frac{b_{ij}}{1 + b_{ij}} \tag{3.27}$$

Eşitlik (3.17)'den yararlanılarak Eşitlik (3.28) ve Eşitlik (3.29) elde edilebilir.

$$\alpha_{ij} = exp\left[-2\left\{\lambda_{M_1}(1) + \lambda_{Y_1M_1}(i,1) + \lambda_{Y_2M_1}(j,1) + \lambda_{M_1M_2}(1,1)\right\}\right] \tag{3.28}$$

$$b_{ij} = exp\left[-2\left\{\lambda_{M_2}(1) + \lambda_{Y_1M_2}(i,1) + \lambda_{Y_2M_2}(j,1) + \lambda_{M_1M_2}(1,1)\right\}\right] \tag{3.29}$$

Eşitlik (3.30) ile odds oranı elde edilebilir.

$$\theta = \left[4\lambda_{M_1 M_2}(1,1)\right] \tag{3.30}$$

α_{ij} ve b_{ij} değerlerinin her biri i veya j'den yalnızca birine bağlıysa ya da hiçbirine bağlı değilse $\alpha_{ij} \in \{\alpha_{i.}, \alpha_{.j}, \alpha_{..}\}$ ve $b_{ij} \in \{\beta_{i.}, \beta_{.j}, \beta_{..}\}$ olarak kabul edilir.

α_{ij} ve b_{ij} kayıp gözlem türlerine göre aşağıdaki durumlarda incelenebilir (Baker, Rosenberger ve Dersimonian, 1992)

1) $\alpha_{ij} = \alpha_{i.}$ ise Y_1'deki kayıp durumu NMAR
2) $\alpha_{ij} = \alpha_{.j}$ ise Y_1'deki kayıp durumu MAR
3) $\alpha_{ij} = \alpha_{..}$ ise Y_1'deki kayıp durumu MCAR
4) $b_{ij} = \beta_{i.}$ ise Y_2'deki kayıp durumu MAR
5) $b_{ij} = \beta_{.j}$ ise Y_2'deki kayıp durumu NMAR
6) $b_{ij} = \beta_{..}$ ise Y_2'deki kayıp durumu MCAR

Δ, μ_{ij}'lerden bağımsız bir sabit terim olmak üzere, Eşitlik (3.31)'de Poisson örneklemesi altında μ'nün log-olabilirlik fonksiyonu verilmiştir.

$$l(\mu; y_{göz}) = \sum_{i,j} y_{ij11} log\mu_{ij11} + \sum_{i} y_{i+12} log\mu_{i+12} + \sum_{j} y_{+j21} log\mu_{+j21}$$
$$+ y_{++22} log\mu_{++22} - \mu_{++++} + \Delta \tag{3.31}$$

Yukarıda tanımlanan altı duruma göre sekiz tane model tanımlanır. Bu modellerin kapalı form tahminleri aşağıda verilmiştir:

Model 1: $\boldsymbol{\alpha_{..}}, \boldsymbol{\beta_{i.}}$ (Y_1 MCAR ve Y_2 MAR)

Y_1 değişkenindeki kayıp gözlemlerin iki değişkenlerden de bağımsız olduğu, Y_2 değişkenindeki kayıp gözlemlerin Y_1 değişkenine bağlı olduğu durumdur. $i = 1,2, \ldots, R$, $j = 1,2, \ldots, C$ ve $x, s = 1, 2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.32)'deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1 Y_2} + \lambda_{is}^{Y_1 M_2} + \lambda_{xs}^{M_1 M_2} \tag{3.32}$$

Model parametreleri Eşitlik (3.33) ve Eşitlik (3.34) yardımıyla; odds oranı ise Eşitlik (3.35) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{..} = \frac{y_{++21}}{y_{++11}} \tag{3.33}$$

$$\hat{\beta}_{i.} = \frac{y_{i+12}}{\hat{\mu}_{i+}}, \qquad i = 1,2,\dots,R \tag{3.34}$$

$$\hat{\theta} = \frac{y_{++11}y_{++22}}{y_{++12}y_{++11}} \tag{3.35}$$

Beklenen sıklıklar Eşitlik (3.36) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ij11} = \frac{y_{ij11}y_{+j+1}y_{++11}}{y_{+j11}y_{+++1}} \tag{3.36}$$

Model 2: $\boldsymbol{\alpha}_{..}, \boldsymbol{\beta}_{.j}$ (Y$_1$ MCAR ve Y$_2$ NMAR)

Y$_1$ değişkenindeki kayıp gözlemlerin iki değişkenlerden de bağımsız olduğu, Y$_2$ değişkenindeki kayıp gözlemlerin yalnızca Y$_2$'ye yani değişkenin kendisine bağlı olduğu durumdur. $i = 1,2,\dots,R,\ j = 1,2,\dots,C$ ve $x,s\ =\ 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.37)'deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \tag{3.37}$$

$\hat{\alpha}_{..}$ parametresi Eşitlik (3.33), odds oranı Eşitlik (3.35) yardımıyla hesaplanabilir. $\hat{\beta}_{.j}$ ifadesi $\sum_j \hat{\mu}_{ij11}\hat{\beta}_{.j} = y_{i+12}$ eşitliğini sağlar.

Beklenen sıklıklar ise Eşitlik (3.36) yardımıyla hesaplanabilir.

Model 3: $\boldsymbol{\alpha}_{i.}, \boldsymbol{\beta}_{..}$ (Y$_1$ NMAR ve Y$_2$ MCAR)

Y$_1$ değişkenindeki kayıp gözlemlerin yalnızca Y$_1$'e yani değişkenin kendisine bağlı olduğu, Y$_2$ değişkenindeki kayıp gözlemlerin iki değişkenlerden de bağımsız olduğu durumdur. $i = 1,2,\dots,R, j = 1,2,\dots,C$ ve $x,s\ =\ 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.38)'deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{xs}^{M_1M_2} \tag{3.38}$$

$\hat{\alpha}_{i.}$ ifadesi $\sum_i \hat{\mu}_{ij11}\hat{\alpha}_{i.} = y_{+j21}$ eşitliğini sağlar. $\hat{\beta}_{..}$ parametresi Eşitlik (3.39), odds oranı Eşitlik (3.35) yardımıyla hesaplanabilir.

$$\hat{\beta}_{..} = \frac{y_{++12}}{y_{++11}} \tag{3.39}$$

Beklenen sıklıklar Eşitlik (3.40) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ij11} = \frac{y_{ij11}y_{i+1+}y_{++11}}{y_{i+11}y_{+++1}} \tag{3.40}$$

Model 4: $\boldsymbol{\alpha_{i.}}, \boldsymbol{\beta_{i.}}$ (Y1 NMAR ve Y2 MAR)

Y1 değişkenindeki kayıp gözlemlerin yalnızca Y1’e yani değişkenin kendisine bağlı olduğu, Y2 değişkenindeki kayıp gözlemlerin Y1 değişkenine bağlı olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$ ve $x,s = 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.41)’deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{ix}^{Y_1M_2} + \lambda_{xs}^{M_1M_2} \tag{3.41}$$

$\hat{\alpha}_{i..}$ ifadesi $\sum_i \hat{\mu}_{ij11}\hat{\alpha}_{i.} = y_{+j21}$ eşitliğini sağlar. $\hat{\beta}_{i.}$ parametresi Eşitlik (3.42), odds oranı Eşitlik (3.43) yardımıyla hesaplanabilir.

$$\hat{\beta}_{i.} = \frac{y_{i+12}}{y_{i+11}}, \qquad i = 1,2,\dots,R \tag{3.42}$$

$$\hat{\theta} = \frac{y_{++22}}{\sum_i y_{i+}\hat{\alpha}_{i.}\beta_{i.}} \tag{3.43}$$

Beklenen sıklıklar Eşitlik (3.44) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ij11} = y_{ij11} \tag{3.44}$$

Model 5: $\boldsymbol{\alpha_{i.}}, \boldsymbol{\beta_{.j}}$ (Y1 ve Y2 NMAR)

Y1 değişkenindeki kayıp gözlemlerin yalnızca Y1’e, Y2 değişkenindeki kayıp gözlemlerin yalnızca Y2’ye yani değişkenlerin kendisine bağlı olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$ ve $x,s = 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.45)’teki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{is}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \tag{3.45}$$

Odds oranı Eşitlik (3.46) yardımıyla hesaplanabilir.

$$\hat{\theta} = \frac{y_{++22}}{\sum_{i,j} y_{ij}\hat{\alpha}_{i.}\hat{\beta}_{.j}} \tag{3.46}$$

$\hat{\alpha}_{i.}$ ifadesi $\sum_i \hat{\mu}_{ij11}\hat{\alpha}_{i.} = y_{+j21}$ ve $\hat{\beta}_{.j}$ ifadesi $\sum_j \hat{\mu}_{ij11}\hat{\beta}_{.j} = y_{i+12}$ eşitliğini sağlar.

Beklenen sıklıklar Eşitlik (3.44) yardımıyla hesaplanabilir.

Model 6: $\boldsymbol{\alpha_{.j}, \beta_{..}}$ (Y_1 MAR ve Y_2 MCAR)

Y_1 değişkenindeki kayıp gözlemlerin Y_2 değişkenine bağlı olduğu, Y_2 değişkenindeki kayıp gözlemlerin iki değişkenlerden de bağımsız olduğu durumdur. $i = 1,2,\dots,R,\;\; j = 1,2,\dots,C$ ve $x, s \; = \; 1{,}2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.47)'deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{xs}^{M_1M_2} \tag{3.47}$$

$\hat{\alpha}_{.j}$ parametresi Eşitlik (3.48) kullanılarak tahmin edilir. $\hat{\beta}_{..}$ parametresi Eşitlik (3.39), odds oranı Eşitlik (3.35) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{.j} = \frac{y_{+j21}}{\hat{\mu}_{+j}}, \qquad j = 1,2,\dots,C \tag{3.48}$$

Beklenen sıklıklar Eşitlik (3.49) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ij11} = \frac{y_{ij11}y_{i+1+}y_{++11}}{y_{i+11}y_{++1+}} \tag{3.49}$$

Model 7: $\boldsymbol{\alpha_{.j}, \beta_{i.}}$ (Y_1 ve Y_2 MAR)

Y_1 değişkenindeki kayıp gözlemlerin Y_2 değişkenine, Y_2 değişkenindeki kayıp gözlemlerin Y_1 değişkenine bağlı olduğu durumdur. $i = 1,2,\dots,R,\;\; j = 1,2,\dots,C$ ve $x, s \; = \; 1{,}2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.50)'deki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_S^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{is}^{Y_1M_2} + \lambda_{xs}^{M_1M_2} \tag{3.50}$$

$\hat{\beta}_{i.}$ parametresi Eşitlik (3.42) yardımıyla hesaplanabilirken, $\hat{\alpha}_{.j}$ parametresi Eşitlik (3.51) ve odds oranı Eşitlik (3.52) yardımıyla tahmin edilebilir.

$$\hat{\alpha}_{.j} = \frac{y_{+j21}}{\hat{\mu}_{+j11}}, \qquad j = 1,2,\dots,C \tag{3.51}$$

$$\hat{\theta} = \frac{y_{++22}}{\sum_{i,j} y_{ij+11}\hat{\alpha}_{i.}\hat{\beta}_{i.}} \tag{3.52}$$

Beklenen sıklıklar Eşitlik (3.44) yardımıyla hesaplanabilir.

Model 8: $\boldsymbol{\alpha}_{.j}, \boldsymbol{\beta}_{.j}$ (Y_1 MAR ve Y_2 NMAR)

Y_1 değişkenindeki kayıp gözlemlerin Y_2 değişkenine, Y_2değişkenindeki kayıp gözlemlerin yalnızca Y_2'ye yani değişkenin kendisine bağlı olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$ ve $x, s = 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.53)'teki gibi tanımlanır.

$$log\mu_{ijxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{jx}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \tag{3.53}$$

$\hat{\alpha}_{.j}$ ifadesi Eşitlik (3.51) kullanılarak tahmin edilir. Odds oranı Eşitlik (3.54) yardımıyla hesaplanabilir.

$$\hat{\theta} = \frac{y_{++22}}{\sum_j y_{+j}\hat{\alpha}_{.j}\hat{\beta}_{.j}} \tag{3.54}$$

$\hat{\beta}_{.j.}$ ifadesi $\sum_j \hat{\mu}_{ij11}\hat{\beta}_{.j} = y_{i+12}$ eşitliğini sağlar.

Beklenen sıklıklar Eşitlik (3.44) yardımıyla hesaplanabilir.

Yukarıda verilen sekiz model için hipotezler:

H_0: Modele uyum vardır.

H_s: Modele uyum yoktur.

şeklinde kurulur. Bu hipotezleri test edebilmek için olabilirlik oran istatistiği Eşitlik (3.55)'te verilmiştir.

$$\begin{aligned} G^2 = -2\Bigg[&\sum_{i,j} y_{ij11} ln\left(\frac{\hat{\mu}_{ij11}}{\hat{\mu}_{ij11}}\right) + \sum_i y_{i+12} ln\left(\frac{\sum_j \hat{\mu}_{ij11}\hat{b}_{ij}}{y_{i+12}}\right) \\ &+ \sum_j y_{+j21} ln\left(\frac{\sum_i \hat{\mu}_{ij11}\hat{\alpha}_{ij}}{y_{+j21}}\right) + y_{++22} ln\left(\frac{\sum_{i,j} \hat{\mu}_{ij11}\hat{\alpha}_{ij}\hat{b}_{ij}\hat{\theta}}{y_{++22}}\right) \\ &- \sum_{i,j} \hat{\mu}_{ij11}\left(1 + \hat{\alpha}_{ij} + \hat{b}_{ij} + \hat{\alpha}_{ij}\hat{b}_{ij}\hat{\theta}\right) + n\Bigg] \end{aligned} \tag{3.55}$$

Bu sekiz model için serbestlik dereceleri:

$$sd = (R+1)(C+1) - \text{modelde tahmin edilen parametre sayısı}$$

olarak hesaplanır. Modellerde tahmin edilen parametre sayıları Çizelge 3.4'te yer almaktadır.

Çizelge 3. 4. İki boyutlu tablolarda iki değişkende kayıp durumunda modellerden tahmin edilen parametre sayıları

Model	Parametre Sayısı
$(\alpha_{..}, \beta_{i.})$	$RC + R + 2$
$(\alpha_{..}, \beta_{.j})$	$RC + C + 2$
$(\alpha_{i.}, \beta_{..})$	$RC + R + 2$
$(\alpha_{.j}, \beta_{..})$	$RC + C + 2$
$(\alpha_{i.}, \beta_{i.})$	$RC + 2R + 1$
$(\alpha_{.j}, \beta_{.j})$	$RC + 2C + 1$
$(\alpha_{i.}, \beta_{.j})$	$RC + R + C + 1$
$(\alpha_{.j}, \beta_{i.})$	$RC + R + C + 1$

G^2 değeri $X^2_{a;sd}$ değeri ile karşılaştırılır. $G^2 < X^2_{a;sd}$ ise hipotez $\%a$ anlamlılık düzeyinde reddedilemez. Modele uyum vardır.

3.2. Kayıp Gözlem İçeren Üç Boyutlu Olumsallık Tabloları İçin Logaritmik Doğrusal Modeller

Bu bölümde Ghosh ve Vellaisamy (2016, 2019, 2024) çalışmalarında kayıp gözlem içeren üç boyutlu olumsallık tabloları için önerilen logaritmik doğrusal modeller tanıtılmıştır.

Y_1, Y_2 ve Y_3 sırasıyla R, C ve K düzeylerine sahip üç kategorik değişken olmak üzere, tanımlama kolaylığı için R=C=K=2 olarak ele alınmıştır. RxCxK boyutlu olumsallık tablolarında yalnızca tek değişkende (tabaka değişkeninde) kayıp olması ve iki değişkende (hem tabaka hem satır değişkenlerinde) kayıp olması durumları incelenmiştir (Ghosh ve Vellaisamy, 2024).

3.2.1. Tek Değişkende Kayıp Olması Durumu

Y_1 (tabaka) değişkeninde kayıp gözlemler mevcut olsun. Kayıp durumu, Y_1 gözleniyorsa $M = 1$, Y_1 kayıpsa $M = 2$ olmak üzere M ile gösterilsin. Bu durumda olumsallık tablosu Çizelge 3.5'teki gibi olur (Ghosh ve Vellaisamy, 2024).

Çizelge 3. 5. Tek değişkende kayıp olan üç boyutlu olumsallık tablosu

			Y3 = 1	**Y3 = 2**
M = 1	**Y1 = 1**	**Y2 = 1**	y_{1111}	y_{1121}
		Y2 = 2	y_{1211}	y_{1221}
	Y1 = 2	**Y2 = 1**	y_{2111}	y_{2121}
		Y2 = 2	y_{2211}	y_{2221}
M = 2	**Kayıp**	**Y2 = 1**	y_{+112}	y_{+122}
		Y2 = 2	y_{+212}	y_{+222}

Çizelge 3.5'te verilen olumsallık tablosuna ait eşitlikler aşağıdaki gibi tanımlanır (Ghosh ve Vellaisamy, 2024).

$i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x = 1,2$ olmak üzere göze sıklıkları $\{y_{ijkx}\}$ şeklinde gösterilmiştir.

Gözlenen sıklıklar vektörü $y_{göz} = (\{y_{ijk1}\},\{y_{+jk2}\})$ olmak üzere $\{y_{ijk1}\}$ tamamen gözlenen; $\{y_{+jk2}\}$ ise Y_1 değişkeninin kayıp olduğu durumdaki değerlerden oluşmaktadır. Burada gösterilen "+" işareti Y_1 değişkeni kayıp olduğu için tüm durumlardaki toplamı ifade etmektedir.

Örneklem oranları vektörü $\pi = \{\pi_{ijkx}\}$ ve beklenen sıklıklar vektörü $\mu = \{\mu_{ijkx}\}$ şeklinde gösterilir. $n = \sum_{i,j,k,x} y_{ijkx}$ olmak üzere toplam örneklem büyüklüğünü ifade etmektedir. Çizelge 3.3'te verilen olumsallık tablosu için $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x = 1,2$ olmak üzere logaritmik doğrusal model Eşitlik (3.56)'daki gibi tanımlanır.

$$log\mu_{ijkx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} + \lambda_{ix}^{Y_1M} + \lambda_{jx}^{Y_2M} + \lambda_{kx}^{Y_3M} \tag{3.56}$$

Bu modeldeki tüm parametreler, parametre tahminlerinin toplamın sıfır olması koşulunu sağlar. Örneğin; $\sum_i \lambda_{Y_1Y_3(i,k)} = \sum_k \lambda_{Y_1Y_3(i,k)} = 0$

Y_1 değişkenindeki kayıp türünü belirlemek için odds Eşitlik (3.57)'de verilmiştir.

$$\alpha_{ijk} = \frac{P(M = 2 \mid Y_1 = i, Y_2 = j, Y_3 = k)}{P(M = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k)} = \frac{\pi_{ijk2}}{\pi_{ijk1}} = \frac{\mu_{ijk2}}{\mu_{ijk1}} \tag{3.57}$$

Beklenen sıklıklar Eşitlik (3.58)'de verilmiştir.

$$\mu_{ijk2} = \alpha_{ijk}\mu_{ijk1} \tag{3.58}$$

Beklenen sıklıklar Eşitlik (3.59)'da yer alan koşulu sağlar.

$$\sum_{ijk} \mu_{ijk1}\left(1 + \alpha_{ijk}\right) = n \tag{3.59}$$

Birleşik olasılıklar Eşitlik (3.60)'taki gibi hesaplanır ve buradan marjinal olasılıklar elde edilebilir.

$$\pi_{ijk+} = \frac{\mu_{ijk1}\left(1 + \alpha_{ijk}\right)}{n} \tag{3.60}$$

Eşitlik (3.56)'dan yararlanılarak

$$\alpha_{ijk} = exp\left[-2\left\{\lambda_M(1) + \lambda_{Y_1M}(i,1) + \lambda_{Y_2M}(j,1) + \lambda_{Y_3M}(k,1)\right\}\right] \tag{3.61}$$

elde edilebilir.

a_{ijk}, kayıp gözlem türlerine göre aşağıdaki dört durumda incelenebilir (Ghosh ve Vellaisamy,2016).

1) $\alpha_{ijk} = \alpha_{i..}$ ise NMAR türü kayıp gözlem (kayıp durumu yalnızca Y_1'e yani değişkenin kendisine bağlıdır.)
2) $\alpha_{ijk} = \alpha_{.j.}$ ise MAR türü kayıp gözlem (kayıp durumu Y_2 değişkenine bağlıdır.)
3) $\alpha_{ijk} = \alpha_{..k}$ ise MAR türü kayıp gözlem (kayıp durumu Y_3 değişkenine bağlıdır.)
4) $\alpha_{ijk} = \alpha_{...}$ ise MCAR türü kayıp gözlem (kayıp durumu diğer değişkenlerden bağımsızdır.)

Δ bir sabit olmak üzere, Poisson örneklemesi altında μ'nün log-olabilirlik fonksiyonu Eşitlik (3.62) yardımıyla hesaplanabilir.

$$l\left(\mu; y_{göz}\right) = \sum_{i,j,k} y_{ijk1} log\mu_{ijk1} + \sum_{j,k} y_{+jk2} log\mu_{+jk2} - \sum_{i,j,k,x} \mu_{ijkx} + \Delta \tag{3.62}$$

Yukarıda tanımlanan NMAR, MAR ve MCAR türü kayıp gözlemlerin olduğu üç durum için logaritmik doğrusal modeller ve bu modellerin kapalı form tahminleri aşağıda verilmiştir:

Model 1: $\boldsymbol{\alpha_{ijk} = \alpha_{i..}}$ (Y$_1$ NMAR)

Y$_1$ değişkenindeki kayıp gözlemlerin yalnızca Y$_1$'e yani değişkenin kendisine bağlı olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x = 1,2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.63)'te yer almaktadır.

$$log\mu_{ijkx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} + \lambda_{ix}^{Y_1M} \tag{3.63}$$

$j,k \geq 1$ olmak üzere $\hat{\alpha}_{i..}$ parametresi $\sum_i \hat{\mu}_{ijk1}\hat{\alpha}_{i..} = y_{+jk2}$ eşitliğini sağlar.

Beklenen sıklıklar Eşitlik (3.64)'te yer almaktadır.

$$\hat{\mu}_{ijk1} = y_{ijk1} \tag{3.64}$$

Model 2: $\boldsymbol{\alpha_{ijk} = \alpha_{.j.}}$ (Y$_1$ MAR)

Y$_1$ değişkenindeki kayıp gözlemlerin Y$_2$ değişkenine bağlı olarak ortaya çıktığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x = 1,2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.65)'te yer almaktadır.

$$log\mu_{ijkx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} + \lambda_{jx}^{Y_2M} \tag{3.65}$$

Model parametresi Eşitlik (3.66) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{.j.} = \frac{y_{+j+2}}{y_{+j+1}}, \qquad j = 1,2,\dots,C \tag{3.66}$$

Beklenen sıklıklar Eşitlik (3.67)'de yer almaktadır.

$$\hat{\mu}_{ijk1} = \frac{y_{ijk1}y_{+jk+}y_{+j+1}}{y_{+jk1}y_{+j++}} \tag{3.67}$$

Model 3: $\boldsymbol{\alpha_{ijk} = \alpha_{..k}}$ (Y$_1$ MAR)

Y$_1$ değişkenindeki kayıp gözlemlerin Y$_3$ değişkenine bağlı olarak ortaya çıktığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x = 1,2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.68)'de yer almaktadır.

$$log\mu_{ijkx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} + \lambda_{kx}^{Y_3M} \tag{3.68}$$

Model parametresi Eşitlik (3.69) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{..k} = \frac{y_{++k2}}{y_{++k1}}, \qquad k = 1,2,\dots,K \tag{3.69}$$

Beklenen sıklıklar Eşitlik (3.70)'te yer almaktadır.

$$\hat{\mu}_{ijk1} = \frac{y_{ijk1}y_{+jk+}y_{++k1}}{y_{+jk1}y_{++k+}} \tag{3.70}$$

Model 4: $\boldsymbol{\alpha_{ijk} = \alpha_{...}}$ (Y_1 MCAR)

Y_1 değişkenindeki kayıp gözlemlerin diğer tüm değişkenlerden bağımsız olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x = 1,2$ olmak üzere, logaritmik doğrusal model denklemi Eşitlik (3.71)'de yer almaktadır.

$$log\mu_{ijkx} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} \tag{3.71}$$

Model parametresi Eşitlik (3.72) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{...} = \frac{y_{+++2}}{y_{+++1}} \tag{3.72}$$

Beklenen sıklıklar Eşitlik (3.73)'te yer almaktadır.

$$\hat{\mu}_{ijk1} = \frac{y_{ijk1}y_{+jk+}y_{+++1}}{y_{+jk1}y_{++++}} \tag{3.73}$$

Yukarıda verilen dört model için hipotezler:

H_0: Modele uyum vardır.

H_s: Modele uyum yoktur.

şeklinde kurulur. Bu hipotezleri test edebilmek için olabilirlik oran istatistiği Eşitlik (3.74)'te verilmiştir.

$$G^2 = -2\left[\sum_{i,j,k} y_{ijk1} ln\left(\frac{\hat{\mu}_{ijk1}}{y_{ijk1}}\right) + \sum_{j,k} y_{+jk2} ln\left(\frac{\sum_i \hat{\mu}_{ijk1}\hat{\alpha}_{ijk}}{y_{+jk2}}\right) - \sum_{i,j,k} \hat{\mu}_{ijk1}\left(1+\hat{\alpha}_{ijk}\right) + n\right] \tag{3.74}$$

Bu dört model için serbestlik dereceleri:

$$sd = (R+1)CK - \text{modelde tahmin edilen parametre sayısı}$$

olarak hesaplanır. Modellerde tahmin edilen parametre sayıları Çizelge 3.6'da verilmiştir.

Çizelge 3. 6. Üç boyutlu tablolarda tek değişkende kayıp durumunda modellerden tahmin edilen parametre sayıları

Model	**Parametre Sayısı**
$(\alpha_{i..})$	$R(1+CK)$
$(\alpha_{.j.})$	$C(1+RK)$
$(\alpha_{..k})$	$K(1+RC)$
$(\alpha_{...})$	$RCK+1$

G^2 değeri $X_{a;sd}^2$ değeri ile karşılaştırılır. $G^2 < X_{a;sd}^2$ ise hipotez $\%a$ anlamlılık düzeyinde reddedilemez. Modele uyum vardır.

3.2.2. İki Değişkende Kayıp Olması Durumu

Y_1, Y_2 ve Y_3 sırasıyla R, C ve K düzeylerine sahip üç kategorik değişken olmak üzere, tanımlama kolaylığı için R=C=K=2 olarak ele alınmıştır. Y_1 değişkenindeki kayıp durumu, Y_1 gözleniyorsa $M_1 = 1$, Y_2 kayıpsa $M_1 = 2$ olmak üzere M_1 adlında yeni bir değişken ile; Y_2 değişkenindeki kayıp durumu ise, Y_2 gözleniyorsa $M_2 = 1$, Y_2 kayıpsa $M_2 = 2$ olmak üzere M_2 adlında yeni bir değişken ile gösterilsin. Bu durumda oluşacak olumsallık tablosu Çizelge 3.7'deki gibi oluşturulur (Ghosh ve Vellaisamy, 2024).

Çizelge 3. 7. İki değişkende kayıp olan üç boyutlu olumsallık tablosu

				$Y_3 = 1$	**$Y_3 = 2$**
$M_1 = 1$	**$Y_1 = 1$**	**$M_2 = 1$**	$Y_2 = 1$	y_{11111}	y_{11211}
			$Y_2 = 2$	y_{12111}	y_{12211}
		$M_2 = 2$	Kayıp	y_{1+112}	y_{1+212}
	$Y_1 = 2$	**$M_2 = 1$**	$Y_2 = 1$	y_{21111}	y_{21211}
			$Y_2 = 2$	y_{22111}	y_{22211}
		$M_2 = 2$	Kayıp	y_{2+112}	y_{2+212}
$M_1 = 2$	**Kayıp**	**$M_2 = 1$**	$Y_2 = 1$	y_{+1121}	y_{+1221}
			$Y_2 = 2$	y_{+2121}	y_{+2221}
		$M_2 = 2$	Kayıp	y_{++122}	y_{++222}

Çizelge 3.7'de verilen olumsallık tablosuna ait eşitlikler aşağıdaki gibi tanımlanır (Ghosh ve Vellaisamy, 2024).

$i = 1,2,\ldots,R,\ j = 1,2,\ldots,C,\ k = 1,2,\ldots,K$ ve $x,s\ =\ 1,2$ olmak üzere göze sıklıkları $\{y_{ijkxs}\}$ şeklinde gösterilmiştir.

Gözlenen sıklıklar vektörü $y_{göz} = (\{y_{ijk11}\},\{y_{+jk21}\},\{y_{i+k12}\},\{y_{++k22}\})$ olmak üzere $\{y_{ijk11}\}$ tamamen gözlenen, $\{y_{+jk21}\}$ Y_1 değişkeninin kayıp olduğu durumdaki; $\{y_{i+k12}\}$ Y_2 değişkeninin kayıp olduğu durumdaki ve $\{y_{++k22}\}$ ise hem Y_1 hem de Y_2 değişkenlerinin kayıp olduğu durumdaki değerlerden oluşmaktadır. Burada gösterilen "+" işareti kayıp değişkenlerde tüm durumlardaki toplamı ifade etmektedir.

Örneklem oranları vektörü; $\pi = \{\pi_{ijkxs}\}$ ve beklenen sıklıklar vektörü $\mu = \{\mu_{ijkxs}\}$ şeklinde gösterilir. n ise örneklem büyüklüğüdür. $ii = 1,2,\ldots,R,\ \ j = 1,2,\ldots,C,$ $k = 1,2,\ldots,K$ ve $x,s\ =\ 1,2$ olmak üzere bu olumsallık tablosu için logaritmik doğrusal model Eşitlik (3.75)'teki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ik}^{Y_1Y_3} + \lambda_{jk}^{Y_2Y_3} + \lambda_{ix}^{Y_1M_1} + \lambda_{jx}^{Y_2M_1}$$
$$+ \lambda_{kx}^{Y_3M_1} + \lambda_{is}^{Y_1M_2} + \lambda_{js}^{Y_2M_2} + \lambda_{ks}^{Y_3M_2} + \lambda_{xs}^{M_1M_2} \tag{3.75}$$

Bu modeldeki tüm parametreler, parametre tahminlerinin toplamın sıfır olması koşulunu sağlar. Örneğin, $\sum_i \lambda_{Y_1Y_2(i,j)} = \sum_j \lambda_{Y_1Y_2(i,j)} = 0$

Y_1 ve Y_2 değişkenlerindeki kayıp gözlemler için oddslar sırasıyla α_{ijk} ve b_{ijk} şeklinde gösterilir.

Y_2 gözlendiğinde, Y_1 değişkeninde kayıp olması durumu için odds Eşitlik (3.76) yardımıyla hesaplanabilir.

$$\alpha_{ijk} = \frac{P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k)}{P(M_1 = 1, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k)} = \frac{\pi_{ijk21}}{\pi_{ijk11}} = \frac{\mu_{ijk21}}{\mu_{ijk11}} \tag{3.76}$$

Y_1 gözlendiğinde, Y_2 değişkeninde kayıp olması durumu için odds Eşitlik (3.77) yardımıyla hesaplanabilir.

$$b_{ijk} = \frac{P(M_1 = 1, M_2 = 2 \mid Y_1 = i, Y_2 = j, Y_3 = k)}{P(M_1 = 1, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k)} = \frac{\pi_{ijk12}}{\pi_{ijk11}} = \frac{\mu_{ijk12}}{\mu_{ijk11}} \tag{3.77}$$

M_1 ve M_2 arasındaki odds oranı (θ) Eşitlik (3.78) yardımıyla hesaplanabilir.

$$\theta = \frac{P(M_1 = 1, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k)P(M_1 = 2, M_2 = 2 \mid Y_1 = i, Y_2 = j, Y_3 = k)}{P(M_1 = 1, M_2 = 2 \mid Y_1 = i, Y_2 = j, Y_3 = k)P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k)}$$

$$= \frac{\pi_{ijk11}\pi_{ijk22}}{\pi_{ijk12}\pi_{ijk21}} = \frac{\mu_{ijk11}\mu_{ijk22}}{\mu_{ijk12}\mu_{ijk21}} \quad (3.78)$$

Eğer $\theta = 1$ ise M_1 ve M_2 bağımsızdır.

Beklenen sıklıklar Eşitlik (3.79) – Eşitlik (3.81) yardımıyla hesaplanabilir.

$$\mu_{ijk21} = \alpha_{ijk}\mu_{ijk11} \quad (3.79)$$

$$\mu_{ijk12} = b_{ijk}\mu_{ijk11} \quad (3.80)$$

$$\mu_{ijk22} = \mu_{ijk11}\alpha_{ijk}b_{ijk}\theta \quad (3.81)$$

Toplam örneklem büyüklüğü Eşitlik (3.82) kullanılarak hesaplanabilir.

$$n = \sum_{i,j,k} \mu_{ijk11}\left(1 + \alpha_{ijk} + b_{ijk} + \alpha_{ijk}b_{ijk}\theta\right) \quad (3.82)$$

Örneklem oranları Eşitlik (3.83) kullanılarak hesaplanır.

$$\pi_{ijk++} = \frac{\mu_{ijk11}\left(1 + \alpha_{ijk} + b_{ijk} + \alpha_{ijk}b_{ijk}\theta\right)}{n} \quad (3.83)$$

Y_2 gözlenirken Y_1 değişkeninde kayıp olması olasılığı Eşitlik (3.84) ile, Y_1 gözlenirken Y_2 değişkeninde kayıp olması olasılığı Eşitlik (3.85) ile hesaplanabilir.

$$\phi_{1|2}(i,j,k) = P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k) = \frac{\alpha_{ijk}}{1 + \alpha_{ijk}} \quad (3.84)$$

$$\phi_{2|1}(i,j,k) = P(M_1 = 2, M_2 = 1 \mid Y_1 = i, Y_2 = j, Y_3 = k) = \frac{b_{ijk}}{1 + b_{ijk}} \quad (3.85)$$

Eşitlik (3.75)'ten yararlanılarak Eşitlik (3.86) ve Eşitlik (3.87) elde edilebilir.

$$\alpha_{ijk} = exp\left[-2\left\{\lambda_{M_1}(1) + \lambda_{Y_1M_1}(i,1) + \lambda_{Y_2M_1}(j,1) + \lambda_{Y_3M_1}(k,1) + \lambda_{M_1M_2}(1,1)\right\}\right] \quad (3.86)$$

$$b_{ijk} = exp\left[-2\left\{\lambda_{M_2}(1) + \lambda_{Y_1M_2}(i,1) + \lambda_{Y_2M_2}(j,1) + \lambda_{Y_3M_2}(k,1) + \lambda_{M_1M_2}(1,1)\right\}\right] \quad (3.87)$$

Eşitlik (3.88) ile odds oranı elde edilebilir.

$$\theta = exp\left[4\lambda_{M_1M_2}(1,1)\right] \tag{3.88}$$

α_{ijk} ve b_{ijk} değerlerinin her biri i, j veya k'dan yalnızca birine bağlıysa ya da hiçbirine bağlı değilse $\alpha_{ijk} \in \{\alpha_{i..}, \alpha_{.j.}, \alpha_{..k}, \alpha_{...}\}$ ve $b_{ijk} \in \{\beta_{i..}, \beta_{.j.}, \beta_{..k}, \beta_{...}\}$ olarak kabul edilir.

α_{ijk} ve b_{ijk} kayıp gözlem türlerine göre aşağıdaki durumlarda incelenebilir (Ghosh ve Vellaisamy, 2024):

1) $\alpha_{ijk} = \alpha_{i..}$ ise Y_1'deki kayıp durumu NMAR
2) $\alpha_{ijk} = \alpha_{.j.}$ veya $\alpha_{ijk} = \alpha_{..k}$ ise Y_1'deki kayıp durumu MAR
3) $\alpha_{ijk} = \alpha_{...}$ ise Y_1'deki kayıp durumu MCAR
4) $b_{ijk} = \beta_{.j.}$ ise Y_2'deki kayıp durumu NMAR
5) $b_{ijk} = \beta_{i..}$ veya $b_{ijk} = \beta_{..k}$ ise Y_2'deki kayıp durumu MAR
6) $b_{ijk} = \beta_{...}$ ise Y_2'deki kayıp durumu MCAR

Eşitlik (3.89)'da Poisson örneklemesi altında μ'nün log-olabilirlik fonksiyonu verilmiştir.

$$l(\mu; y_{göz}) = \sum_{i,j,k} y_{ijk11} log\mu_{ijk11} + \sum_{j,k} y_{+jk21} log\mu_{+jk21} + \sum_{i,k} y_{i+k12} log\mu_{i+k12}$$

$$+ \textstyle\sum_k y_{++k22} log\mu_{++k12} - \sum_{ijkxs} \mu_{ijkxs} \tag{3.89}$$

Yukarıda verilen durumlara göre tanımlanabilen on beş adet logaritmik doğrusal model vardır. Bu modeller ve modellere ait kapalı form tahminleri aşağıda verilmiştir.

Model 1: $\boldsymbol{\alpha_{...}}, \boldsymbol{\beta_{i..}}$ (Y_1 MCAR ve Y_2 MAR)

Y_1 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu, Y_2 değişkenindeki kayıp gözlemlerin Y_1'den kaynaklandığı durumdur. $i = 1,2, \dots, R$, $j = 1,2, \dots, C$, $k = 1,2, \dots, K$ ve $x, s = 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.90)'daki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{is}^{Y_1M_2} + \lambda_{xs}^{M_1M_2} \tag{3.90}$$

Model parametreleri Eşitlik (3.91) ve Eşitlik (3.92) yardımıyla; odds oranı ise Eşitlik (3.93) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{...} = \frac{y_{+++21}}{y_{+++11}} \tag{3.91}$$

$$\hat{\beta}_{i..} = \frac{y_{i++12}}{\hat{\mu}_{i++11}}, \qquad i = 1,2,\dots,R \tag{3.92}$$

$$\hat{\theta} = \frac{y_{+++11}y_{+++22}}{y_{+++12}y_{+++21}} \tag{3.93}$$

Beklenen sıklıklar Eşitlik (3.94) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ijk11} = \frac{y_{ijk11}y_{+++11}y_{+jk+1}}{y_{++++1}y_{+jk11}} \tag{3.94}$$

Model 2: $\boldsymbol{\alpha}_{...}, \boldsymbol{\beta}_{.j.}$ (Y1 MCAR ve Y2 NMAR)

Y_1 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu, Y_2 değişkenindeki kayıp gözlemlerin yalnızca Y_2'ye yani değişkenin kendisine bağlı olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x, s = 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.95)'teki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{js}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \tag{3.95}$$

$\hat{\alpha}_{...}$ parametresi Eşitlik (3.91), odds oranı Eşitlik (3.93) yardımıyla hesaplanabilir. $\hat{\beta}_{.j.}$ ifadesi $\sum_j \hat{\mu}_{ijk11}\hat{\beta}_{.j.} = y_{i+k12}$ eşitliğini sağlar.

Beklenen sıklıklar ise Eşitlik (3.94) yardımıyla hesaplanabilir.

Model 3: $\boldsymbol{\alpha}_{...}, \boldsymbol{\beta}_{..k}$ (Y1 MCAR ve Y2 MAR)

Y_1 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu, Y_2 değişkenindeki kayıp gözlemlerin Y_3'ten kaynaklandığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x, s = 1,2$ olmak üzere, logaritmik doğrusal model Eşitlik (3.96)'daki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ks}^{Y_3M_2} + \lambda_{xs}^{M_1M_2} \tag{3.96}$$

$\hat{\alpha}_{...}$ parametresi Eşitlik (3.91), $\hat{\beta}_{..k}$ parametresi Eşitlik (3.97), odds oranı Eşitlik (3.93) ve beklenen sıklıklar Eşitlik (3.94) yardımıyla hesaplanabilir.

$$\hat{\beta}_{...k} = \frac{y_{++k12}}{\hat{\mu}_{++k11}}, \qquad k = 1,2, \dots, K \tag{3.97}$$

Model 4: $\boldsymbol{\alpha_{i..}}, \boldsymbol{\beta_{...}}$ (Y_1 NMAR ve Y_2 MCAR)

Y_1 değişkenindeki kayıp gözlemlerin yalnızca Y_1'e yani değişkenin kendisine bağlı olduğu, Y_2 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu durumdur. $i = 1,2, \dots, R,\; j = 1,2, \dots, C,\; k = 1,2, \dots, K$ ve $x, s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.98)'deki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{xs}^{M_1M_2} \tag{3.98}$$

$\hat{\alpha}_{i..}$ ifadesi $\sum_i \hat{\mu}_{ijk11}\hat{\alpha}_{i..} = y_{+jk21}$ eşitliğini sağlar. $\hat{\beta}_{...}$ parametresi Eşitlik (3.99), odds oranı Eşitlik (3.93) yardımıyla hesaplanabilir.

$$\hat{\beta}_{...} = \frac{y_{+++12}}{y_{+++11}} \tag{3.99}$$

Beklenen sıklıklar ise Eşitlik (3.100) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ijk11} = \frac{y_{ijk11}y_{+++11}y_{i+k1+}}{y_{+++1+}y_{i+k11}} \tag{3.100}$$

Model 5: $\boldsymbol{\alpha_{i..}}, \boldsymbol{\beta_{i..}}$ (Y_1 NMAR ve Y_2 MAR)

Y_1 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu, Y_2 değişkenindeki kayıp gözlemlerin Y_1'den kaynaklandığı durumdur. $i = 1,2, \dots, R$, $j = 1,2, \dots, C$, $k = 1,2, \dots, K$ ve $x, s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.101)'deki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{is}^{Y_1M_2} + \lambda_{xs}^{M_1M_2} \tag{3.101}$$

$\hat{\alpha}_{i..}$ ifadesi $\sum_i \hat{\mu}_{ijk11}\hat{\alpha}_{i..} = y_{+jk21}$ eşitliğini sağlar. $\hat{\beta}_{i..}$ parametresi Eşitlik (3.102), odds oranı Eşitlik (3.103) yardımıyla hesaplanabilir.

$$\hat{\beta}_{i..} = \frac{y_{i++12}}{y_{i++11}}, \qquad i = 1,2, \dots, R \tag{3.102}$$

$$\hat{\theta} = \frac{y_{+++22}}{\sum_i y_{i++12}\hat{\alpha}_{i..}} \qquad (3.103)$$

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

$$\hat{\mu}_{ijk11} = y_{ijk11} \qquad (3.104)$$

Model 6: $\boldsymbol{\alpha_{i..}}, \boldsymbol{\beta_{.j.}}$ (Y1 ve Y2 NMAR)

Y_1 ve Y_2 değişkenlerindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu durumdur. $i = 1,2,\dots,R,\ \ j = 1,2,\dots,C,\ k = 1,2,\dots,K$ ve $x,s\ =\ 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.105)'teki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{js}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \qquad (3.105)$$

Odds oranı Eşitlik (3.106) yardımıyla hesaplanabilir.

$$\hat{\theta} = \frac{y_{+++22}}{\sum_{i,j} y_{ij+11}\hat{\alpha}_{i..}\hat{\beta}_{.j.}} \qquad (3.106)$$

$\hat{\alpha}_{i..}$ ifadesi $\sum_i \hat{\mu}_{ijk11}\hat{\alpha}_{i..} = y_{+jk21}$ ve $\hat{\beta}_{.j.}$ ifadesi $\sum_j \hat{\mu}_{ijk11}\hat{\beta}_{.j.} = y_{i+k12}$ eşitliğini sağlar.

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 7: $\boldsymbol{\alpha_{i..}}, \boldsymbol{\beta_{..k}}$ (Y1 NMAR ve Y2 MAR)

Y_1 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu, Y_2 değişkenindeki kayıp gözlemlerin Y_3'ten kaynaklandığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C,\ k = 1,2,\dots,K$ ve $x,s\ =\ 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.107)'deki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{ix}^{Y_1M_1} + \lambda_{kx}^{Y_3M_2} + \lambda_{xs}^{M_1M_2} \qquad (3.107)$$

$\hat{\alpha}_{i..}$ ifadesi $\sum_i \hat{\mu}_{ijk11}\hat{\alpha}_{i..} = y_{+jk21}$ eşitliğini sağlar. $\hat{\beta}_{..k}$ parametresi Eşitlik (3.108), odds oranı Eşitlik (3.109) yardımıyla hesaplanabilir.

$$\hat{\beta}_{..k} = \frac{y_{++k12}}{y_{++k11}}, \qquad k = 1,2,\dots,K \qquad (3.108)$$

$$\hat{\theta} = \frac{y_{+++22}}{\sum_{i,k} y_{i+k11}\hat{\alpha}_{i..}\hat{\beta}_{..k}} \qquad (3.109)$$

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 8: $\boldsymbol{\alpha}_{.j.}, \boldsymbol{\beta}_{...}$ (Y_1 MAR ve Y_2 MCAR)

Y_1 değişkenindeki kayıp durumunun Y_2'den kaynaklandığı, Y_2 değişkenindeki kayıp gözlemlerin diğer tüm değişkenlerden bağımsız olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x,s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.110)'daki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{xs}^{M_1M_2} \qquad (3.110)$$

$\hat{\alpha}_{.j.}$ parametresi Eşitlik (3.111) kullanılarak tahmin edilir. $\hat{\beta}_{...}$ parametresi Eşitlik (3.99), odds oranı Eşitlik (3.93) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{.j.} = \frac{y_{+j+21}}{\hat{\mu}_{+j+11}}, \qquad j = 1,2,\dots,C \qquad (3.111)$$

Beklenen sıklıklar ise Eşitlik (3.100) yardımıyla hesaplanabilir.

Model 9: $\boldsymbol{\alpha}_{.j.}, \boldsymbol{\beta}_{i..}$ (Y_1 ve Y_2 MAR)

Y_1 ve Y_2'nin MAR türü kayıp gözleme sahip olduğu; Y_1'deki kayıp gözlemlerin Y_2'den, Y_2'deki kayıp gözlemlerin ise Y_1'den kaynaklandığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x,s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.112)'deki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{is}^{Y_1M_2} + \lambda_{xs}^{M_1M_2} \quad (3.112)$$

$\hat{\alpha}_{.j.}$ parametresi Eşitlik (3.113) ve $\hat{\beta}_{i..}$ parametresi Eşitlik (3.102) yardımıyla hesaplanabilirken, odds oranı Eşitlik (3.114) yardımıyla tahmin edilebilir.

$$\hat{\alpha}_{.j.} = \frac{y_{+j+21}}{y_{+j+11}} \qquad (3.113)$$

$$\hat{\theta} = \frac{y_{+++22}}{\sum_{i,j} y_{ij+11}\hat{\alpha}_{.j.}\hat{\beta}_{i..}} \qquad (3.114)$$

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 10: $\boldsymbol{\alpha}_{.j.}, \boldsymbol{\beta}_{.j.}$ (Y1 MAR ve Y2 NMAR)

Y_1 değişkenindeki kayıp gözlemlerin Y_2'den kaynaklandığı, Y_2 değişkenindeki kayıp gözlemlerin diğer değişkenlerden bağımsız olduğu durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x,s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.115)'teki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{js}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \quad (3.115)$$

$\hat{\alpha}_{.j.}$ ifadesi Eşitlik (3.113) kullanılarak tahmin edilir. Odds oranı Eşitlik (3.116) yardımıyla hesaplanabilir.

$$\hat{\theta} = \frac{y_{+++22}}{\sum_j y_{+j+21}\hat{\beta}_{.j.}} \quad (3.116)$$

$\hat{\beta}_{.j.}$ ifadesi $\sum_j \hat{\mu}_{ijk11}\hat{\beta}_{.j.} = y_{i+k12}$ eşitliğini sağlar.

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 11: $\boldsymbol{\alpha}_{.j.}, \boldsymbol{\beta}_{..k}$ (Y1 ve Y2 MAR)

Y_1 ve Y_2'nin MAR türü kayıp gözleme sahip olduğu; Y_1'deki kayıp gözlemlerin Y_2'den, Y_2'deki kayıp gözlemlerin ise Y_3'ten kaynaklandığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x,s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.117)'deki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{jx}^{Y_2M_1} + \lambda_{ks}^{Y_3M_2} + \lambda_{xs}^{M_1M_2} \quad (3.117)$$

$\hat{\alpha}_{.j.}$ parametresi Eşitlik (3.113) ve $\hat{\beta}_{..k}$ parametresi Eşitlik (3.108) yardımıyla hesaplanabilirken, odds oranı Eşitlik (3.118) yardımıyla tahmin edilebilir.

$$\hat{\theta} = \frac{y_{+++22}}{\sum_{j,k} y_{+jk11}\hat{\alpha}_{.j.}\hat{\beta}_{..k}} \quad (3.118)$$

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 12: $\boldsymbol{\alpha}_{..k}, \boldsymbol{\beta}_{...}$ (Y1 MAR ve Y2 MCAR)

Y_1'deki kayıp gözlemlerin Y_3'ten kaynaklandığı, Y_2 değişkenindeki kayıp gözlemlerin diğer tüm değişkenlerden bağımsız olduğu durumdur.

$i = 1,2, \dots, R,\ \ j = 1,2, \dots, C,\ \ k = 1,2, \dots, K$ ve $x, s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.119)'daki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{kx}^{Y_3M_1} + \lambda_{xs}^{M_1M_2} \qquad (3.119)$$

$\hat{\alpha}_{..k}$ parametresi Eşitlik (3.120) kullanılarak tahmin edilir. $\hat{\beta}_{...}$ parametresi Eşitlik (3.99), odds oranı Eşitlik (3.93) yardımıyla hesaplanabilir.

$$\hat{\alpha}_{..k} = \frac{y_{++k21}}{\hat{\mu}_{++k11}}, \qquad k = 1,2, \dots, K \qquad (3.120)$$

Beklenen sıklıklar ise Eşitlik (3.100) yardımıyla hesaplanabilir.

Model 13: $\boldsymbol{\alpha}_{..k}, \boldsymbol{\beta}_{i..}$ (Y1 ve Y2 MAR)

Y_1 ve Y_2'nin MAR türü kayıp gözleme sahip olduğu; Y_1'deki kayıp gözlemlerin Y_3'ten, Y_2'deki kayıp gözlemlerin ise Y_1'den kaynaklandığı durumdur. $i = 1,2, \dots, R$, $j = 1,2, \dots, C$, $k = 1,2, \dots, K$ ve $x, s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.121)'deki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{kx}^{Y_3M_1} + \lambda_{is}^{Y_1M_2} + \lambda_{xs}^{M_1M_2} \qquad (3.121)$$

$\hat{\alpha}_{..k}$ parametresi Eşitlik (3.122) ve $\hat{\beta}_{i..}$ parametresi Eşitlik (3.102) yardımıyla hesaplanabilirken, odds oranı Eşitlik (3.123) yardımıyla tahmin edilebilir.

$$\hat{\alpha}_{..k} = \frac{y_{++k21}}{y_{++k11}} \qquad (3.122)$$

$$\hat{\theta} = \frac{y_{+++22}}{\sum_{i,k} y_{i+k11}\hat{\alpha}_{..k}\hat{\beta}_{i..}} \qquad (3.123)$$

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 14: $\boldsymbol{\alpha}_{..k}, \boldsymbol{\beta}_{.j.}$ (Y1 MAR ve Y2 NMAR)

Y_1 değişkenindeki kayıp gözlemlerin Y_3'ten kaynaklandığı, Y_2 değişkenindeki kayıp durumunun diğer değişkenlerden bağımsız olduğu durumdur. $i = 1,2, \dots, R$, $j = 1,2, \dots, C$, $k = 1,2, \dots, K$ ve $x, s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.124)'teki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{kx}^{Y_3M_1} + \lambda_{js}^{Y_2M_2} + \lambda_{xs}^{M_1M_2} \qquad (3.124)$$

$\hat{\alpha}_{..k}$ ifadesi Eşitlik (3.122) kullanılarak tahmin edilir. Odds oranı Eşitlik (3.125) yardımıyla hesaplanabilir.

$$\hat{\theta} = \frac{y_{+++22}}{\sum_{j,k} y_{+jk11}\hat{\alpha}_{..k}\hat{\beta}_{.j.}} \tag{3.125}$$

$\hat{\beta}_{.j.}$ ifadesi $\sum_j \hat{\mu}_{ijk11}\hat{\beta}_{.j.} = y_{i+k12}$ eşitliğini sağlar.

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Model 15: $\boldsymbol{\alpha}_{..k}, \boldsymbol{\beta}_{..k}$ (Y_1 ve Y_2 MAR)

Y_1 ve Y_2'nin MAR türü kayıp gözleme sahip olduğu, hem Y_1 hem de Y_2'deki kayıp gözlemlerin Y_3'ten kaynaklandığı durumdur. $i = 1,2,\dots,R$, $j = 1,2,\dots,C$, $k = 1,2,\dots,K$ ve $x, s = 1,2$ olmak üzere olmak üzere, logaritmik doğrusal model Eşitlik (3.126)'daki gibi tanımlanır.

$$log\mu_{ijkxs} = \lambda + \lambda_i^{Y_1} + \lambda_j^{Y_2} + \lambda_k^{Y_3} + \lambda_x^{M_1} + \lambda_s^{M_2} + \lambda_{ij}^{Y_1Y_2} + \lambda_{kx}^{Y_3M_1} + \lambda_{ks}^{Y_3M_2} + \lambda_{xs}^{M_1M_2} \tag{3.126}$$

$\hat{\alpha}_{..k}$ parametresi Eşitlik (3.122) ve $\hat{\beta}_{..k}$ parametresi Eşitlik (3.108) yardımıyla hesaplanabilirken, odds oranı Eşitlik (3.127) yardımıyla tahmin edilebilir.

$$\hat{\theta} = \frac{y_{+++22}}{\sum_k y_{i+k12}\hat{\alpha}_{..k}} \tag{3.127}$$

Beklenen sıklıklar ise Eşitlik (3.104) yardımıyla hesaplanabilir.

Yukarıda verilen on beş model için hipotezler:

H_0: Modele uyum vardır.

H_s: Modele uyum yoktur.

şeklinde kurulur. Bu hipotezleri test edebilmek için olabilirlik oran istatistiği Eşitlik (3.128)'de verilmiştir.

$$G^2 = -2\left[\sum_{i,j,k} y_{ijk11} ln\left(\frac{\hat{\mu}_{ijk11}}{y_{ijk11}}\right) + \sum_{j,k} y_{+jk21} ln\left(\frac{\sum_i \hat{\mu}_{ijk11}\hat{\alpha}_{ijk}}{y_{+jk21}}\right)\right.$$

$$+ \sum_{i,k} y_{i+k12} ln\left(\frac{\sum_j \hat{\mu}_{ijk11}\hat{b}_{ijk}}{y_{i+k12}}\right) + \sum_{k} y_{++k22} ln\left(\frac{\sum_{i,j} \hat{\mu}_{ijk11}\hat{\alpha}_{ijk}\hat{b}_{ijk}\hat{\theta}}{y_{++k22}}\right)$$

$$\left. - \sum_{i,j,k} \hat{\mu}_{ijk11}\left(1 + \hat{\alpha}_{ijk} + \hat{b}_{ijk} + \hat{\alpha}_{ijk}\hat{b}_{ijk}\hat{\theta}\right) + n\right] \quad (3.128)$$

Bu on beş model için serbestlik dereceleri:

$sd \; = \; (R+1)(C+1)K \; -$ modelde tahmin edilen parametre sayısı

olarak hesaplanır. Modellerde tahmin edilen parametre sayıları Çizelge 3.8'de verilmiştir.

Çizelge 3. 8. Üç boyutlu tablolarda iki değişkende kayıp durumunda modellerden tahmin edilen parametre sayıları

Model	**Parametre Sayısı**
$(\alpha_{...}, \beta_{i..})$	$RCK + R + 2$
$(\alpha_{...}, \beta_{.j.})$	$RCK + C + 2$
$(\alpha_{...}, \beta_{..k})$	$RCK + K + 2$
$(\alpha_{i..}, \beta_{...})$	$RCK + R + 2$
$(\alpha_{i..}, \beta_{i..})$	$RCK + 2R + 1$
$(\alpha_{i..}, \beta_{.j.})$	$RCK + R + C + 1$
$(\alpha_{i..}, \beta_{..k})$	$RCK + R + K + 1$
$(\alpha_{.j.}, \beta_{...})$	$RCK + C + 2$
$(\alpha_{.j.}, \beta_{i..})$	$RCK + R + C + 1$
$(\alpha_{.j.}, \beta_{.j.})$	$RCK + 2C + 1$
$(\alpha_{.j.}, \beta_{..k})$	$RCK + C + K + 1$
$(\alpha_{..k}, \beta_{...})$	$RCK + C + 2$
$(\alpha_{..k}, \beta_{i..})$	$RCK + R + K + 1$
$(\alpha_{..k}, \beta_{.j.})$	$RCK + C + K + 1$
$(\alpha_{..k}, \beta_{..k})$	$RCK + 2K + 1$

G^2 değeri $X^2_{a;sd}$ değeri ile karşılaştırılır. $G^2 < X^2_{a;sd}$ ise hipotez $\%a$ anlamlılık düzeyinde reddedilemez. Modele uyum vardır.

4. UYGULAMA

Bu bölümde, üçüncü bölümde tanıtılan kayıp gözlem içeren iki ve üç boyutlu olumsallık tablolarının çözümlenmesinde kullanılan logaritmik doğrusal modellerin gerçek bir veri seti üzerinden uygulaması yapılmıştır. Sağlık çalışmalarında hastaların bazı bilgileri paylaşmak istememesi ya da eksik paylaşması, kontrole gelmemesi, takibinin ya da tedavinin duraklatılması gibi farklı nedenlerden kaynaklanan kayıp gözlem sorunu ile sıklıkla karşılaşılmaktadır.

Bu çalışmada da uygulama için bir sağlık verisi tercih edilmiştir. Kaliforniya Üniversitesi Irvine (UCI) veri deposundan (UC Irvine Machine Learning Repository) ulaşılan farklı değişkenlerde kayıp gözlemlerin mevcut olduğu “Miyokard Enfarktüsü Komplikasyonları” (Golovenkin ve ark., 2020) veri seti kullanılmıştır. Bu veri setinde 1700 hastaya ait veriler bulunmaktadır. Değişkenler içerisinden kalp krizi (1: kalp krizi var, 0: kalp krizi yok), hipertansiyon (1: hipertansiyon var, 0: hipertansiyon yok) ve cinsiyet (1: kadın, 2: erkek) değişkenleri dikkate alınarak olumsallık tabloları oluşturulmuştur.

4.1. İki Boyutlu Olumsallık Tablosu İçin Modeller

“Miyokard Enfarktüsü Komplikasyonları” veri setinden kalp krizi ve hipertansiyon değişkenleri kullanılarak Çizelge 4.1’de yer alan olumsallık tablosu oluşturulmuştur. Bazı hastaların kalp krizi, bazı hastaların hipertansiyon, bazı hastaların ise kalp krizi ve hipertansiyon verilerinin ikisinde de kayıplar olduğu tespit edilmiştir.

Çizelge 4. 1. Kalp Krizi x Hipertansiyon olumsallık tablosu

Kalp Krizi	**Hipertansiyon**		
	Var	Yok	Kayıp
Var	446	187	3
Yok	640	416	4
Kayıp	0	2	2

Kalp krizi ve hipertansiyon değişkenleri arasında ilişki olup olmadığını incelemek amacıyla yalnızca gözlenen sıklıklar kullanılarak süreklilik düzeltmeli Ki-Kare testi uygulanmıştır. Test sonuçları ve odds oranı Çizelge 4.2.’de özetlenmiştir.

Çizelge 4. 2. Kalp Krizi x Hipertansiyon tablosunu Ki-Kare testi sonuçları ve odds oranı

X^2	sd	p	θ	θ için %95 Güven Aralığı
16,308	1	<0,001	1,550	[1,256; 1,914]

Kalp Krizi ve hipertansiyon değişkenleri arasında önemli bir ilişki olduğu %5 anlamlılık düzeyinde söylenebilir ($p<0,001$; Çizelge 4.2). Hipertansiyonu olan birinde olmayan birine göre kalp krizi görülme riskinin 1,55 kat daha fazla olduğu söylenebilir. Odds oranının %95 güven aralığı sınırları [1,256; 1,914] içinde "1" değeri yer almadığı için hesaplanan odds oranının %5 anlamlılık düzeyinde önemli olduğu söylenebilir.

4.1.1. Tek Değişkende Kayıp Olması Durumu

Kalp Krizi x Hipertansiyon olumsallık tablosundan yalnızca hipertansiyon değişkenindeki kayıp gözlemler dikkate alınarak oluşturulan alt tablo Çizelge 4.3'te verilmiştir.

Çizelge 4. 3. Hipertansiyonda kayıp gözlem bulunan Kalp Krizi x Hipertansiyon tablosu

Kalp Krizi	**Hipertansiyon**		
	Var	Yok	Kayıp
Var	446	187	3
Yok	640	416	4

Kalp Krizi x Hipertansiyon tablosunda kayıp gözlem oranı 0,004 (7/1696) olarak elde edilir. İki boyutlu tablolarda sadece sütun değişkeninde kayıp olması durumunda kullanılması önerilen ve Eşitlik (3.8), Eşitlik (3.10) ve Eşitlik (3.13)'te verilen üç model Çizelge 4.3'te yer alan veriye uygulanmış ve model sonuçları Çizelge 4.4'te özetlenmiştir.

Çizelge 4. 4. Hipertansiyonda kayıp gözlem bulunan Kalp Krizi x Hipertansiyon tablosu için model sonuçları

Model	**Parametre Tahminleri**	**G^2**	**sd**	**p**
$(b_{i.})$	$\hat{\beta}_{1.} = 0,0047$ $\hat{\beta}_{2.} = 0,0038$	0	0	-
$(b_{.j})$	$\hat{\beta}_{.1} = 0,0076$ $\hat{\beta}_{.2} = -0,0021$	0,0003	0	-
$(b_{..})$	$\hat{\beta}_{..} = 0,0041$	0,085	1	0,770

Çizelge 4.4'te yer alan $(b_{..})$ (Y_1 MCAR) modeline uyum bulunmuştur ($G^2 = 0{,}085$; $sd = 1$; $p = 0{,}770$). Bu modele göre hipertansiyon değişkenindeki kayıp durumu tamamen bağımsızdır. $(b_{..})$ modeli için beklenen sıklıklar Çizelge 4.5'te özetlenmiştir.

Çizelge 4. 5. Model $(b_{..})$ için beklenen sıklıklar

Kalp Krizi	**Hipertansiyon**			
	Gözlenen		**Kayıp**	
	Var	Yok	Var	Yok
Var	446,2642	187,1108	1,8495	0,7754
Yok	640,3791	416,2464	2,6540	1,7251

Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu Çizelge 4.6'da verilmiştir.

Çizelge 4. 6. Hipertansiyonda kayıp gözlem bulunan Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu

Kalp Krizi	**Hipertansiyon**	
	Var	**Yok**
Var	448,1137	187,8862
Yok	643,0331	417,9715

Beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta} = \frac{448{,}1137 \times 417{,}9715}{187{,}8862 \times 643{,}0331} = 1{,}55$$

Çalışmanın sonucunda, hipertansiyonu olan birinde olmayan birine göre kalp krizi görülme riskinin 1,55 kat daha fazla olduğu söylenebilir.

4.1.2. İki Değişkende Kayıp Olması Durumu

Kalp Krizi x Hipertansiyon olumsallık tablosunda hem kalp krizi hem de hipertansiyon değişkenindeki kayıp gözlemler dikkate alınarak oluşturulan tablo Çizelge 4.1'de verilmişti.

Modellere ait G^2 değerlerinin hesaplanabilmesi için kayıp gözlemin mevcut olmadığı hücrelerde "0" olan değerler "2" ile değiştirilmiştir (Ghosh ve Vellaisamy, 2024).

İki boyutlu tablolarda iki değişkende de kayıp olması durumunda kullanılması önerilen ve Eşitlik (3.32), Eşitlik (3.37), Eşitlik (3.38), Eşitlik (3.41), Eşitlik (3.45), Eşitlik (3.47), Eşitlik (3.50) ve Eşitlik (3.53)'te verilen sekiz model Çizelge 4.1'de yer alan veriye uygulanmış ve model sonuçları Çizelge 4.7'de özetlenmiştir.

Çizelge 4. 7. Kalp krizi ve hipertansiyonda kayıp gözlem bulunan Kalp Krizi x Hipertansiyon tablosu için model sonuçları

Model	**Parametre Tahminleri**	**G^2**	**sd**	**p**	**AIC**	**BIC**
$(\alpha_{..}, \beta_{i.})$	$\hat{\alpha}_{..} = 0{,}0024$ $\hat{\beta}_{1.} = 0{,}0047$ $\hat{\beta}_{2.} = 0{,}0038$ $\hat{\theta} = 120{,}6429$	0,340	1	0,560	-1,660	-7,100
$(\alpha_{..}, \beta_{.j})$	$\hat{\alpha}_{..} = 0{,}0024$ $\hat{\beta}_{.1} = 0{,}0076$ $\hat{\beta}_{.2} = -0{,}0021$ $\hat{\theta} = 120{,}6429$	0,340	1	0,560	-1,660	-7,100
$(\alpha_{i.}, \beta_{..})$	$\hat{\alpha}_{1.} = -0{,}0068$ $\hat{\alpha}_{2.} = 0{,}0079$ $\hat{\beta}_{..} = 0{,}0041$ $\hat{\theta} = 120{,}6429$	2,094	1	0,148	0,094	-5,346
$(\alpha_{.j}, \beta_{..})$	$\hat{\alpha}_{.1} = 0{,}0018$ $\hat{\alpha}_{.2} = 0{,}0033$ $\hat{\beta}_{..} = 0{,}0041$ $\hat{\theta} = 120{,}6429$	0,085	1	0,771	**-1,915**	**-7,355**
$(\alpha_{i.}, \beta_{i.})$	$\hat{\alpha}_{1.} = 0{,}0079$ $\hat{\alpha}_{2.} = -0{,}0068$ $\hat{\beta}_{1.} = 0{,}0038$ $\hat{\beta}_{2.} = 0{,}0047$ $\hat{\theta} = 178{,}5714$	0	0	-		
$(\alpha_{.j}, \beta_{.j})$	$\hat{\alpha}_{.1} = 0{,}0033$ $\hat{\alpha}_{.2} = 0{,}0018$ $\hat{\beta}_{.1} = -0{,}0021$ $\hat{\beta}_{.2} = 0{,}0076$ $\hat{\theta} = 181{,}8182$	0	0	-		
$(\alpha_{i.}, \beta_{.j})$	$\hat{\alpha}_{1.} = 0{,}0079$ $\hat{\alpha}_{2.} = -0{,}0068$ $\hat{\beta}_{.1} = -0{,}0021$ $\hat{\beta}_{.2} = 0{,}0076$ $\hat{\theta} = 179{,}4489$	0	0	-		
$(\alpha_{.j}, \beta_{i.})$	$\hat{\alpha}_{.1} = 0{,}0033$ $\hat{\alpha}_{.2} = 0{,}0018$ $\hat{\beta}_{1.} = 0{,}0038$ $\hat{\beta}_{2.} = 0{,}0047$ $\hat{\theta} = 121{,}0424$	0	0	-		

En küçük AIC veya BIC değerine sahip $\left(\alpha_{.j}, \beta_{..}\right)$ (Y_1 MAR, Y_2 MCAR) modeli en iyi model olarak elde edilir ($G^2 = 0{,}085;\ sd = 1;\ p = 0{,}771$). Bu modele göre kalp krizi değişkenindeki kayıp durumu hipertansiyon değişkenine bağlıyken, hipertansiyon değişkenindeki kayıp durumu tamamen bağımsızdır. En iyi model için beklenen sıklıklar Çizelge 4.8'de yer almaktadır.

Çizelge 4. 8. Model $(\alpha_{.j}, \beta_{..})$ için beklenen sıklıklar

Kalp Krizi		**Hipertansiyon**			
		Gözlenen		**Kayıp**	
		Var	Yok	Var	Yok
Gözlenen	Var	446,2642	187,1107	1,8495	0,7755
	Yok	639,7727	415,8522	2,6515	1,7235
Kayıp	Var	0,8218	0,6206	0,4109	0,3103
	Yok	1,1782	1,3794	0,5891	0,6897

Hem kalp krizi hem de hipertansiyonda kayıp olduğu durumda Kalp Krizi X Hipertansiyon tablosunun beklenen sıklıklar özet tablosu Çizelge 4.9'da verilmiştir.

Çizelge 4. 9. Kalp krizi ve hipertansiyonda kayıp gözlem bulunan Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu

Kalp Krizi	**Hipertansiyon**	
	Var	**Yok**
Var	449,3465	188,8172
Yok	644,1915	419,6448

Beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta} = \frac{449{,}3465 \times 419{,}6448}{188{,}8172 \times 644{,}1915} = 1{,}55$$

Hipertansiyonu olan bireylerde olmayan bireylere göre kalp krizi görülme riskinin 1,55 kat daha fazla olduğu söylenebilir.

4.2. Üç Boyutlu Olumsallık Tablosu İçin Modeller

Bu bölümde uygulama için "Miyokard Enfarktüsü Komplikasyonları" veri setinden hipertansiyon, cinsiyet, kalp krizi değişkenleri kullanılarak Çizelge 4.10'da yer alan olumsallık tablosu oluşturulmuştur.

Çizelge 4. 10. Hipertansiyon x Cinsiyet x Kalp Krizi olumsallık tablosu

Hipertansiyon	**Cinsiyet**	**Kalp Krizi**	
		Var	Yok
Var	Kadın	197	308
	Erkek	249	332
	Kayıp	0	0
Yok	Kadın	26	99
	Erkek	161	317
	Kayıp	0	0
Kayıp	Kadın	2	2
	Erkek	1	2
	Kayıp	0	0

Hipertansiyon x Cinsiyet x Kalp Krizi olumsallık tablosu için yalnızca gözlenen sıklıklar kullanılarak Çizelge 2.2'de yer alan bağımsızlık (M0) ve karşılıklı bağımsızlık (M7) modelleri uygulanmış ve model sonuçları Çizelge 4.11'de verilmiştir.

Çizelge 4. 11. Hipertansiyon x Cinsiyet x Kalp Krizi tablosu için M0 ve M7 model sonuçları

Model	**G^2**	**sd**	**p**
Bağımsızlık modeli (M0)	142,347	4	<0,001
Karşılıklı bağımsızlık modeli (M7)	3,550	1	0,060

Bağımsızlık modeline uyum olmadığı %5 anlamlılık düzeyinde söylenebilir ($p<0,001$; Çizelge 4.11). Hipertansiyon, cinsiyet ve kalp krizi değişkenlerinin birbirinden bağımsız olmadığı söylenebilir.

Karşılıklı bağımsızlık modeline uyum olduğu %5 anlamlılık düzeyinde söylenebilir ($p=0,060$; Çizelge 4.11). Üç değişkenin de birbiri ile ilişkili olduğu söylenebilir.

4.2.1. Tek Değişkende Kayıp Olması Durumu

Hipertansiyon x Cinsiyet x Kalp Krizi olumsallık tablosundan yalnızca hipertansiyon değişkenindeki kayıp gözlemler dikkate alınarak oluşturulan alt tablo Çizelge 4.12'de gösterilmiştir.

Çizelge 4. 12. Hipertansiyonda kayıp gözlem bulunan Hipertansiyon x Cinsiyet x Kalp Krizi olumsallık tablosu

Hipertansiyon	Cinsiyet	Kalp Krizi	
		Var	Yok
Var	Kadın	197	308
	Erkek	249	332
Yok	Kadın	26	99
	Erkek	161	317
Kayıp	Kadın	2	2
	Erkek	1	2

Üç boyutlu tablolarda sadece sütun değişkeninde kayıp olması durumunda kullanılması önerilen ve Eşitlik (3.63), Eşitlik (3.65) ve Eşitlik (3.68)'de verilen üç model Çizelge 4.28'de yer alan veriye uygulanmış ve model sonuçları Çizelge 4.13'te özetlenmiştir.

Çizelge 4. 13. Hipertansiyonda kayıp gözlem bulunan Hipertansiyon x Cinsiyet x Kalp Krizi tablosu için model sonuçları

Model	Parametre Tahminleri	G^2	sd	p	AIC	BIC
$(\alpha_{i..})$	$\hat{\alpha}_{1..} = 0{,}0091$ $\hat{\alpha}_{2..} = -0{,}0047$	0	2	1,000	-	-
$(\alpha_{.j.})$	$\hat{\alpha}_{.1.} = 0{,}0047$ $\hat{\alpha}_{.2.} = 0{,}0038$	1,411	2	0,494	-2,589	-13,475
$(\alpha_{..k})$	$\hat{\alpha}_{..1} = 0{,}0063$ $\hat{\alpha}_{..2} = 0{,}0028$	0,398	2	0,819	**-3,602**	**-14,488**
$(\alpha_{...})$	$\hat{\alpha}_{...} = 0{,}0041$	1,501	3	0,682	-0,499	-5,942

En küçük AIC veya BIC değerine sahip $(\alpha_{..k})$ (MAR) modeli en iyi model olarak elde edilir ($G^2 = 0{,}398$; $sd = 2$; $p = 0{,}819$). Bu modele göre hipertansiyon değişkenindeki kayıp durumu kalp krizi değişkenine bağlıdır.

$(\alpha_{.j.})$, $(\alpha_{..k})$ ve $(\alpha_{...})$ modelleri için elde edilen beklenen sıklıklar sırasıyla Çizelge 4.14, 4.15 ve 4.16'da verilmiştir.

Çizelge 4. 14. Model $(\alpha_{.j.})$ için beklenen sıklıklar

Hipertansiyon		Cinsiyet	Kalp Krizi	
			Var	Yok
Gözlenen	Var	Kadın	197,8248	307,2935
		Erkek	249,2731	331,7700
	Yok	Kadın	26,1088	98,7729
		Erkek	161,1765	316,7804
Kayıp	Var	Kadın	0,9420	0,1243
		Erkek	1,4633	0,4703
	Yok	Kadın	0,9415	0,6088
		Erkek	1,2531	1,1965

Çizelge 4. 15. Model $(\alpha_{..k})$ için beklenen sıklıklar

Hipertansiyon		Cinsiyet	Kalp Krizi	
			Var	Yok
Gözlenen	Var	Kadın	197,5187	307,8821
		Erkek	248,6434	332,0797
	Yok	Kadın	26,0685	98,9621
		Erkek	160,7694	317,0761
Kayıp	Var	Kadın	1,2481	0,8747
		Erkek	1,5712	0,9434
	Yok	Kadın	0,1647	0,2811
		Erkek	1,0159	0,9008

Çizelge 4. 16. Model $(\alpha_{...})$ için beklenen sıklıklar

Hipertansiyon		Cinsiyet	Kalp Krizi	
			Var	Yok
Gözlenen	Var	Kadın	197,9464	307,4824
		Erkek	249,1819	331,6486
	Yok	Kadın	26,1249	98,8336
		Erkek	161,1176	316,6645
Kayıp	Var	Kadın	0,8204	1,2743
		Erkek	1,0327	1,3745
	Yok	Kadın	0,1083	0,4096
		Erkek	0,6677	1,3124

Hipertansiyon x Cinsiyet x Kalp Krizi tablosunun beklenen sıklıklar özet tablosu Çizelge 4.17’de verilmiştir.

Çizelge 4. 17. Hipertansiyonda kayıp gözlem bulunan Hipertansiyon x Cinsiyet x Kalp Krizi tablosunun beklenen sıklıklar özet tablosu

Hipertansiyon	**Cinsiyet**	**Kalp Krizi**	
		Var	Yok
Var	Kadın	198,7668	308,7568
	Erkek	250,2146	333,0231
Yok	Kadın	26,2332	99,2432
	Erkek	161,7853	317,9769

Hipertansiyon değişkeni dikkate alındığında beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta}_{hip(+)} = \frac{308{,}7568 \times 250{,}2146}{198{,}7668 \times 333{,}0231} = 1{,}17 \qquad \hat{\theta}_{hip(-)} = \frac{99{,}2432 \times 161{,}7853}{26{,}2332 \times 317{,}9769} = 1{,}93$$

Hipertansiyonu olan erkeklerde kalp krizi görülme riski hipertansiyonu olan kadınlara göre 1,17 kat daha fazla iken; hipertansiyonu olmayan erkeklerde kalp krizi görülme riski hipertansiyonu olmayan kadınlara göre 1,93 kat daha fazladır.

Cinsiyet değişkeni dikkate alındığında beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta}_{kadın} = \frac{198{,}7668 \times 99{,}2432}{308{,}7568 \times 26{,}2332} = 2{,}43 \qquad \hat{\theta}_{erkek} = \frac{250{,}2146 \times 317{,}9769}{333{,}0231 \times 161{,}7853} = 1{,}47$$

Hipertansiyonu olan kadınlarda kalp krizi görülme riski hipertansiyonu olmayanlara göre 2,43 kat daha fazla iken; hipertansiyonu olan erkeklerde kalp krizi görülme riski hipertansiyonu olmayanlara göre 1,47 kat daha fazladır.

4.2.2. İki Değişkende Kayıp Olması Durumu

Hipertansiyon x Cinsiyet x Kalp Krizi olumsallık tablosunda hipertansiyon ve cinsiyet değişkenlerinde kayıp gözlemler dikkate alınarak oluşturulan tablo Çizelge 4.10’da verilmişti.

Modellere ait G^2 değerlerinin hesaplanabilmesi için kayıp gözlemin mevcut olmadığı hücrelerde “0” olan değerler “2” ile değiştirilmiştir (Ghosh ve Vellaisamy, 2024).

Üç boyutlu tablolarda iki değişkende kayıp olması durumunda kullanılması önerilen ve Eşitlik (3.90) – Eşitlik (3.126)'da verilen on beş model Çizelge 4.10'da yer alan veriye uygulanmış ve model sonuçları Çizelge 4.18'de özetlenmiştir.

Çizelge 4. 18. Kalp krizi ve hipertansiyonda kayıp gözlem bulunan Hipertansiyon x Cinsiyet x Kalp Krizi tablosu için model sonuçları

Model	**Parametre Tahminleri**	**G^2**	**sd**	**p**	**AIC**	**BIC**
$(\alpha_{...}, \beta_{i..})$	$\hat{\alpha}_{...} = 0,00414$ $\hat{\beta}_{1..} = 0,00368$ $\hat{\beta}_{2..} = 0,00663$ $\hat{\theta} = 120,643$	2,592	6	0,858	-9,411	-42,070
$(\alpha_{...}, \beta_{.j.})$	$\hat{\alpha}_{...} = 0,0041$ $\hat{\beta}_{.1.} \cong 0$ $\hat{\beta}_{.2.} \cong 0$ $\hat{\theta} = 120,643$	1,516	6	0,958	-10,484	-43,142
$(\alpha_{...}, \beta_{..k})$	$\hat{\alpha}_{...} = 0,0041$ $\hat{\beta}_{..1} = 0,0063$ $\hat{\beta}_{..2} = 0,0038$ $\hat{\theta} = 120,643$	2,437	6	0,875	-9,563	-42,222
$(\alpha_{i..}, \beta_{...})$	$\hat{\alpha}_{1..} \cong 0$ $\hat{\alpha}_{2..} \cong 0$ $\hat{\beta}_{...} = 0,0047$ $\hat{\theta} = 120,643$	1,509	6	0,959	**-10,491**	**-43,149**
$(\alpha_{i..}, \beta_{i..})$	$\hat{\alpha}_{1..} = 0,0071$ $\hat{\alpha}_{2..} = 0,0003$ $\hat{\beta}_{1..} = 0,0037$ $\hat{\beta}_{2..} = 0,0066$ $\hat{\theta} = 135,135$	2,476	5	0,780	-7,523	-34,739
$(\alpha_{i..}, \beta_{.j.})$	$\hat{\alpha}_{1..} = 0,0071$ $\hat{\alpha}_{2..} = 0,0003$ $\hat{\beta}_{.1.} = -0,0005$ $\hat{\beta}_{.2.} = 0,0085$ $\hat{\theta} = 116,04$	3,241	5	0,663	-6,759	-33,974
$(\alpha_{i..}, \beta_{..k})$	$\hat{\alpha}_{1..} = 0,0071$ $\hat{\alpha}_{2..} = 0,0003$ $\hat{\beta}_{..1} = 0,0063$ $\hat{\beta}_{..2} = 0,0038$ $\hat{\theta} = 102,399$	0,151	5	0,999	-9,849	-37,064
$(\alpha_{.j.}, \beta_{...})$	$\hat{\alpha}_{.1.} = 0,0064$ $\hat{\alpha}_{.2.} = 0,0028$ $\hat{\beta}_{...} = 0,0047$ $\hat{\theta} = 120,643$	2,106	5	0,910	-7,894	-35,110

Çizelge 4. 18. Kalp krizi ve hipertansiyonda kayıp gözlem bulunan Hipertansiyon x Cinsiyet x Kalp Krizi tablosu için model sonuçları (*devam ediyor)*

Model	Parametre Tahminleri	G²	sd	p	AIC	BIC
$(\alpha_{.j.}, \beta_{i..})$	$\hat{\alpha}_{.1.} = 0{,}0064$ $\hat{\alpha}_{.2.} = 0{,}0028$ $\hat{\beta}_{1..} = 0{,}0037$ $\hat{\beta}_{2..} = 0{,}0066$ $\hat{\theta} = 124{,}536$	1,510	5	0,912	-8,490	-35,705
$(\alpha_{.j.}, \beta_{.j.})$	$\hat{\alpha}_{.1.} = 0{,}0064$ $\hat{\alpha}_{.2.} = 0{,}0028$ $\hat{\beta}_{.1.} = -0{,}0005$ $\hat{\beta}_{.2.} = 0{,}0085$ $\hat{\theta} = 170{,}213$	3,580	5	0,611	-6,420	-33,636
$(\alpha_{.j.}, \beta_{..k})$	$\hat{\alpha}_{.1.} = 0{,}0064$ $\hat{\alpha}_{.2.} = 0{,}0028$ $\hat{\beta}_{..1} = 0{,}0063$ $\hat{\beta}_{..2} = 0{,}0038$ $\hat{\theta} = 121{,}069$	1,312	5	0,934	-8,688	-35,904
$(\alpha_{..k}, \beta_{...})$	$\hat{\alpha}_{..1} = 0{,}0047$ $\hat{\alpha}_{..2} = 0{,}0038$ $\hat{\beta}_{...} = 0{,}0047$ $\hat{\theta} = 120{,}643$	2,952	6	0,815	-9,048	-41,707
$(\alpha_{..k}, \beta_{i..})$	$\hat{\alpha}_{..1} = 0{,}0047$ $\hat{\alpha}_{..2} = 0{,}0038$ $\hat{\beta}_{1..} = 0{,}0037$ $\hat{\beta}_{2..} = 0{,}0066$ $\hat{\theta} = 121{,}042$	2,313	5	0,804	-7,687	-34,903
$(\alpha_{..k}, \beta_{.j.})$	$\hat{\alpha}_{..1} = 0{,}0047$ $\hat{\alpha}_{..2} = 0{,}0038$ $\hat{\beta}_{.1.} = -0{,}0005$ $\hat{\beta}_{.2.} = 0{,}0085$ $\hat{\theta} = 110{,}763$	2,864	5	0,721	-7,136	-34,351
$(\alpha_{..k}, \beta_{..k})$	$\hat{\alpha}_{..1} = 0{,}0047$ $\hat{\alpha}_{..2} = 0{,}0038$ $\hat{\beta}_{..1} = 0{,}0063$ $\hat{\beta}_{..2} = 0{,}0038$ $\hat{\theta} = 117{,}272$	2,405	5	0,791	-7,594	-34,810

En küçük AIC veya BIC değerine sahip $(\alpha_{i..}, \beta_{...})$ (Y_1 NMAR, Y_2 MCAR) modeli en iyi model olarak elde edilir ($G^2 = 1{,}509;\ sd = 6;\ p = 0{,}959$). Bu modele göre hipertansiyon değişkenindeki kayıp durumu hipertansiyon değişkeninin kendisine bağlıyken, cinsiyet değişkenindeki kayıp durumu tüm değişkenlerden bağımsızdır.

$(\alpha_{i..}, \beta_{...})$ modeli için elde edilen beklenen sıklıklar Çizelge 4.19'da verilmiştir.

Çizelge 4. 19. Model $(\alpha_{i..}, \beta_{...})$ için beklenen sıklıklar

Hipertansiyon		**Cinsiyet**		**Kalp Krizi**	
				Var	Yok
Gözlenen	Var	Gözlenen	Kadın	196,9505	307,5059
			Erkek	248,9374	331,4675
		Kayıp	Kadın	0,9328	1,4565
			Erkek	1,1791	1,5700
	Yok	Gözlenen	Kadın	26,1542	99,0070
			Erkek	161,9548	317,0225
		Kayıp	Kadın	0,1238	0,4689
			Erkek	0,7671	1,5016
Kayıp	Var	Gözlenen	Kadın	0	0
			Erkek	0	0
		Kayıp	Kadın	0	0
			Erkek	0	0
	Yok	Gözlenen	Kadın	0	0
			Erkek	0	0
		Kayıp	Kadın	0	0
			Erkek	0	0

Hem kalp krizi hem de hipertansiyonda kayıp olduğu durumda Hipertansiyon x Cinsiyet x Kalp Krizi tablosunun beklenen sıklıklar özet tablosu Çizelge 4.20'de verilmiştir.

Çizelge 4. 20. Kalp krizi ve hipertansiyonda kayıp gözlem bulunan Hipertansiyon x Cinsiyet x Kalp Krizi tablosunun beklenen sıklıklar özet tablosu

Hipertansiyon	**Cinsiyet**	**Kalp Krizi**	
		Var	**Yok**
Var	**Kadın**	197,8834	308,9625
	Erkek	250,1166	333,0375
Yok	**Kadın**	26,2780	99,4759
	Erkek	162,7219	318,5240

Hipertansiyon değişkeni dikkate alındığında beklenen sıklıklardan elde edilen odds oranları:

$$\hat{\theta}_{hip(+)} = \frac{308,9625 \times 250,1166}{197,8834 \times 333,0375} = 1,17 \qquad \hat{\theta}_{hip(-)} = \frac{99,4759 \times 162,7219}{26,2780 \times 318,5240} = 1,93$$

Hipertansiyonu olan erkeklerde kalp krizi görülme riski hipertansiyonu olan kadınlara göre 1,17 kat daha fazla iken; hipertansiyonu olmayan erkeklerde kalp krizi görülme riski hipertansiyonu olmayan kadınlara göre 1,93 kat daha yüksektir.

Cinsiyet değişkeni dikkate alındığında beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta}_{kadın} = \frac{197,8834 \times 99,4759}{308,9625 \times 26,2780} = 2,43 \qquad \hat{\theta}_{erkek} = \frac{250,1166 \times 318,5240}{333,0375 \times 162,7219} = 1,47$$

Hipertansiyonu olan kadınlarda kalp krizi görülme riski hipertansiyonu olmayanlara göre 2,43 kat daha fazla iken; hipertansiyonu olan erkeklerde kalp krizi görülme riski hipertansiyonu olmayanlara göre 1,47 kat daha fazladır.

4.3. Farklı Kayıp Gözlem Oranlarının Karşılaştırılması

Kayıp gözlemler göz ardı edildiğinde elde edilen odds oranı ile logaritmik doğrusal model uygulandıktan sonra elde edilen odds oranı aynı sonucu vermiştir. Bu durumun verideki kayıp gözlem sayısının toplam gözlem sayısına göre oldukça az olmasından kaynaklanıyor olabilir. Bu nedenle hipertansiyonda kayıp gözlem bulunan kalp krizi x hipertansiyon olumsallık tablosunda kayıp gözlem oranı arttırıldığında sonucun değişip değişmeyeceğini incelemek amacıyla verideki kayıp gözlem oranı %5, %10, %15 ve %20 olacak şekilde hipotetik veri oluşturulmuş ve modeller incelenmiştir.

4.3.1. Hipertansiyonda Kayıp Gözlem Oranının %5 Olması Durumu

Kalp Krizi x Hipertansiyon olumsallık tablosundan hipertansiyon değişkenindeki kayıp gözlem oranı %5 olacak şekilde hipotetik veri oluşturmak için gerekli hesaplamalar yapılmıştır.

Kayıp gözlem oranının %5 olması için gereken kayıp gözlem sayısı:

$$kayıp\ gözlem\ sayısı = \frac{(1689 + kayıp\ gözlem\ sayısı) \times 5}{100} \cong 89$$

olarak hesaplanmıştır. Hesaplama sonucunda elde edilen 89 kayıp gözlem, Çizelge 4.3'te yer alan gerçek verideki kayıp gözlem orantılarına (3k ve 4k) uygun olarak dağıtılmış ve oluşturulan yeni tablo Çizelge 4.21'de gösterilmiştir.

Çizelge 4. 21. Hipertansiyonda %5 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu

Kalp Krizi	**Hipertansiyon**		
	Var	Yok	Kayıp
Var	446	187	38
Yok	640	416	51

Eşitlik (3.8), Eşitlik (3.10) ve Eşitlik (3.13)'te verilen üç model Çizelge 4.7'de yer alan veriye uygulanmış ve model sonuçları Çizelge 4.22'de özetlenmiştir.

Çizelge 4. 22. Hipertansiyonda %5 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu için model sonuçları

Model	**Parametre Tahminleri**	**G^2**	**sd**	**p**
$(b_{i.})$	$\hat{\beta}_{1.} = 0{,}0600$ $\hat{\beta}_{2.} = 0{,}0483$	0	0	-
$(b_{.j})$	$\hat{\beta}_{.1} = 0{,}0952$ $\hat{\beta}_{.2} = -0{,}0024$	1,806	0	-
$(b_{..})$	$\hat{\beta}_{..} = 0{,}0527$	1,103	1	0,293

Çizelge 4.22'de yer alan $(b_{..})$ (Y_1 MCAR) modeline uyum bulunmuştur ($G^2 = 1{,}103$; $sd = 1$; $p = 0{,}293$). Bu modele göre hipertansiyon değişkenindeki kayıp durumu tamamen bağımsızdır. $(b_{..})$ modeli için beklenen sıklıklar Çizelge 4.23'te yer almaktadır.

Çizelge 4. 23. Hipertansiyonda %5 kayıp gözlem olması durumunda model $(b_{..})$ için beklenen sıklıklar

Kalp Krizi	**Hipertansiyon**			
	Gözlenen		**Kayıp**	
	Var	Yok	Var	Yok
Var	449,1088	188,3034	23,6653	9,92244
Yok	644,4610	418,8997	33,9591	22,0734

Hipertansiyonda %5 kayıp gözlem olması durumunda kalp krizi x hipertansiyon tablosunun beklenen sıklıklar özet tablosu Çizelge 4.24'te verilmiştir.

Çizelge 4. 24. Hipertansiyonda %5 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu

Kalp Krizi	**Hipertansiyon**	
	Var	**Yok**
Var	472,7741	198,2259
Yok	678,4202	440,9731

Beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta} = \frac{472{,}7741 \times 440{,}97314}{198{,}2259 \times 678{,}4202} = 1{,}55$$

4.3.2. Hipertansiyonda Kayıp Gözlem Oranının %10 Olması Durumu

Kalp Krizi x Hipertansiyon olumsallık tablosundan hipertansiyon değişkenindeki kayıp gözlem oranı %10 olacak şekilde hipotetik veri oluşturmak için gerekli hesaplamalar yapılmıştır.

Kayıp gözlem oranının %10 olması için gereken kayıp gözlem sayısı:

$$kayıp\ gözlem\ sayısı = \frac{(1689 + kayıp\ gözlem\ sayısı) \times 10}{100} \cong 188$$

olarak hesaplanmıştır. Hesaplama sonucunda elde edilen 188 kayıp gözlem, Çizelge 4.3’te yer alan gerçek verideki kayıp gözlem orantılarına (3k ve 4k) uygun olarak dağıtılmış ve oluşturulan yeni tablo Çizelge 4.25’te gösterilmiştir.

Çizelge 4. 25. Hipertansiyonda %10 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu

Kalp Krizi	**Hipertansiyon**		
	Var	Yok	Kayıp
Var	446	187	81
Yok	640	416	107

Eşitlik (3.8), Eşitlik (3.10) ve Eşitlik (3.13)’te verilen üç model Çizelge 4.25’te yer alan veriye uygulanmış ve model sonuçları Çizelge 4.26’da özetlenmiştir.

Çizelge 4. 26. Hipertansiyonda %10 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu için model sonuçları

Model	**Parametre Tahminleri**	**G^2**	**sd**	**p**
$(b_{i.})$	$\hat{\beta}_{1.} = 0{,}1279$ $\hat{\beta}_{2.} = 0{,}1013$	0	0	-
$(b_{.j})$	$\hat{\beta}_{.1} = 0{,}2078$ $\hat{\beta}_{.2} = -0{,}0625$	0	0	-
$(b_{..})$	$\hat{\beta}_{..} = 0{,}1113$	2,896	1	0,089

Çizelge 4.26'da yer alan $(b_{..})$ (Y_1 MCAR) modeline uyum bulunmuştur ($G^2 = 2{,}896$; $sd = 1$; $p = 0{,}089$). Bu modele göre hipertansiyon değişkenindeki kayıp durumu tamamen bağımsızdır. $(b_{..})$ modeli için beklenen sıklıklar Çizelge 4.27'de yer almaktadır.

Çizelge 4. 27. Hipertansiyonda %10 kayıp gözlem olması durumunda model $(b_{..})$ için beklenen sıklıklar

Kalp Krizi	**Hipertansiyon**			
	Gözlenen		**Kayıp**	
	Var	Yok	Var	Yok
Var	452,6835	189,8023	50,3875	21,1266
Yok	649,5907	422,2340	72,3049	46,9982

Hipertansiyonda %10 kayıp gözlem olması durumunda kalp krizi x hipertansiyon tablosunun beklenen sıklıklar özet tablosu Çizelge 4.28'de verilmiştir.

Çizelge 4. 28. Hipertansiyonda %10 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu

Kalp Krizi	**Hipertansiyon**	
	Var	**Yok**
Var	503,0711	210,9289
Yok	721,8957	469,2322

Beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta} = \frac{503{,}0711 \times 469{,}2322}{210{,}9289 \times 721{,}8957} = 1{,}55$$

4.3.3. Hipertansiyonda Kayıp Gözlem Oranının %15 Olması Durumu

Kalp Krizi x Hipertansiyon olumsallık tablosundan hipertansiyon değişkenindeki kayıp gözlem oranı %15 olacak şekilde hipotetik veri oluşturmak için gerekli hesaplamalar yapılmıştır.

Kayıp gözlem oranının %15 olması için gereken kayıp gözlem sayısı:

$$kayıp\ gözlem\ sayısı = \frac{(1689 + kayıp\ gözlem\ sayısı) \times 15}{100} \cong 298$$

olarak hesaplanmıştır. Hesaplama sonucunda elde edilen 298 kayıp gözlem, Çizelge 4.3’te yer alan gerçek verideki kayıp gözlem orantılarına (3k ve 4k) uygun olarak dağıtılmış ve oluşturulan yeni tablo Çizelge 4.29’da gösterilmiştir.

Çizelge 4. 29. Hipertansiyonda %15 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu

Kalp Krizi	Hipertansiyon		
	Var	Yok	Kayıp
Var	446	187	128
Yok	640	416	170

Eşitlik (3.8), Eşitlik (3.10) ve Eşitlik (3.13)’te verilen üç model Çizelge 4.29’da yer alan veriye uygulanmış ve model sonuçları Çizelge 4.30’da özetlenmiştir.

Çizelge 4. 30. Hipertansiyonda %15 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu için model sonuçları

Model	Parametre Tahminleri	G^2	sd	p
$(b_{i.})$	$\hat{\beta}_{1.} = 0{,}2022$ $\hat{\beta}_{2.} = 0{,}1609$	0	0	-
$(b_{.j})$	$\hat{\beta}_{.1} = 0{,}3258$ $\hat{\beta}_{.2} = -0{,}0926$	0	0	-
$(b_{..})$	$\hat{\beta}_{..} = 0{,}1764$	4,688	1	0,030

Uygulanan üç modele de uyum bulunamamıştır. Diğer kayıp gözlem oranları ile karşılaştırabilmek adına $(b_{..})$ modeli için beklenen sıklıklar ve odds oranı hesaplanmıştır. Beklenen sıklıklar Çizelge 4.31’de yer almaktadır.

Çizelge 4. 31. Hipertansiyonda %15 kayıp gözlem olması durumunda model ($b_{..}$) için beklenen sıklıklar

Kalp Krizi	Hipertansiyon			
	Gözlenen		**Kayıp**	
	Var	Yok	Var	Yok
Var	455,7719	191,0972	80,4144	33,7163
Yok	654,0225	425,1146	115,3929	75,0054

Hipertansiyonda %15 kayıp gözlem olması durumunda kalp krizi x hipertansiyon tablosunun beklenen sıklıklar özet tablosu Çizelge 4.32'de verilmiştir.

Çizelge 4. 32. Hipertansiyonda %15 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu

Kalp Krizi	Hipertansiyon	
	Var	**Yok**
Var	536,1864	224,8135
Yok	769,4154	500,1201

Beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta} = \frac{536{,}1864 \times 500{,}1201}{224{,}8135 \times 769{,}4154} = 1{,}55$$

4.3.4. Hipertansiyonda Kayıp Gözlem Oranının %20 Olması Durumu

Kalp Krizi x Hipertansiyon olumsallık tablosundan hipertansiyon değişkenindeki kayıp gözlem oranı %20 olacak şekilde hipotetik veri oluşturmak için gerekli hesaplamalar yapılmıştır.

Kayıp gözlem oranının %20 olması için gereken kayıp gözlem sayısı:

$$kayıp\ gözlem\ sayısı = \frac{(1689 + kayıp\ gözlem\ sayısı) \times 20}{100} \cong 422$$

olarak hesaplanmıştır. Hesaplama sonucunda elde edilen 422 kayıp gözlem, Çizelge 4.3'te yer alan gerçek verideki kayıp gözlem orantılarına (3k ve 4k) uygun olarak dağıtılmış ve oluşturulan yeni tablo Çizelge 4.33'te gösterilmiştir.

Çizelge 4. 33. Hipertansiyonda %20 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu

Kalp Krizi	Hipertansiyon		
	Var	Yok	Kayıp
Var	446	187	181
Yok	640	416	241

Eşitlik (3.8), Eşitlik (3.10) ve Eşitlik (3.13)'te verilen üç model Çizelge 4.33'da yer alan veriye uygulanmış ve model sonuçları Çizelge 4.34'te özetlenmiştir.

Çizelge 4. 34. Hipertansiyonda %20 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosu için model sonuçları

Model	Parametre Tahminleri	G^2	sd	p
$(b_{i.})$	$\hat{\beta}_{1.} = 0{,}2859$ $\hat{\beta}_{2.} = 0{,}2282$	0	0	-
$(b_{.j})$	$\hat{\beta}_{.1} = 0{,}4590$ $\hat{\beta}_{.2} = -0{,}1268$	0	0	-
$(b_{..})$	$\hat{\beta}_{..} = 0{,}2498$	6,915	1	0,009

Uygulanan üç modele de uyum bulunamamıştır. Diğer kayıp gözlem oranları ile karşılaştırabilmek adına $(b_{..})$ modeli için beklenen sıklıklar ve odds oranı hesaplanmıştır. Beklenen sıklıklar Çizelge 4.35'te yer almaktadır.

Çizelge 4. 35. Hipertansiyonda %20 kayıp gözlem olması durumunda model $(b_{..})$ için beklenen sıklıklar

Kalp Krizi	Hipertansiyon			
	Gözlenen		Kayıp	
	Var	Yok	Var	Yok
Var	458,8777	192,3994	114,6515	48,0713
Yok	658,4792	428,0115	164,5223	106,9395

Hipertansiyonda %20 kayıp gözlem olması durumunda kalp krizi x hipertansiyon tablosunun beklenen sıklıklar özet tablosu Çizelge 4.36'da verilmiştir.

Çizelge 4. 36. Hipertansiyonda %20 kayıp gözlem olması durumunda Kalp Krizi x Hipertansiyon tablosunun beklenen sıklıklar özet tablosu

Kalp Krizi	**Hipertansiyon**	
	Var	**Yok**
Var	573,5292	240,4707
Yok	823,0015	534,9510

Beklenen sıklıklardan elde edilen odds oranı:

$$\hat{\theta} = \frac{573{,}5292 \times 534{,}9510}{240{,}4707 \times 823{,}0015} = 1{,}55$$

Çalışmanın sonucunda, hipertansiyonu olan birinde olmayan birine göre kalp krizi görülme riskinin 1,55 kat daha fazla olduğu söylenebilir.

Kayıp gözlem oranlarının %5 ve %10 olduğu olumsallık tablolarına hipertansiyon değişkenindeki kayıp durumunun tamamen bağımsız (MCAR) olduğu modele uyum bulunurken, kayıp gözlem oranının %15 ve %20 olduğu durumda bu modele uyum bulunamamıştır. Modele uyum bulunmadığı durumlar da dahil olmak üzere, farklı beklenen sıklıklar elde edilmiş olsa bile hesaplanan odds oranları tüm durumlarda 1,55 olarak bulunmuştur. Verideki kayıp gözlem oranı ne olursa olsun odds oranlarının birbiri ile aynı çıkmasının sebebi kayıp gözlem türünün ihmal edilebilir kayıp gözlem olmasından kaynaklanmaktadır.

5. SONUÇ VE TARTIŞMA

Kayıp gözlem problemi, veri toplama aşamasında ölçüm cihazının eksik ölçüm yapması, anketin eksik cevaplanması, verilerin yanlışlıkla silinmesi, veri ön işleme sırasında olabilen veri kayıpları vb. nedenlerle ortaya çıkabilen; bilimsel araştırmalarda sonuçların doğruluğunu ve çalışmanın güvenilirliğini etkileyen önemli bir sorundur. İstatistiksel yöntemler ve yazılımlar ise verilerin eksiksiz olduğu varsayımına dayanmaktadır. Bu nedenle istatistiksel veri analizlerinden önce kayıp gözlemlerin tespit edilmesi, nedenlerinin anlaşılması ve kayıp gözlem yapısına uygun yöntemlerle çözülmesi gerekmektedir.

Bu tez çalışması kapsamında kategorik verilerde kayıp gözlem problemi ele alınmış, olumsallık tablolarının yapısı ve logaritmik doğrusal modeller konusunda bilgi verilmiş, literatürde yer alan kayıp gözlem türleri ve çözüm yöntemleri araştırılmış, ihmal edilemeyen kayıp gözlemler için önerilen modelleme yöntemlerinden biri olan logaritmik doğrusal model yaklaşımı tablo yapılarına göre detaylı olarak incelenmiştir. Modeller "Miyokard Enfarktüsü Komplikasyonları" verisine uygulanarak sonuçlar tartışılmıştır.

Uygulamanın birinci bölümünde iki boyutlu olumsallık tablolarında, ikinci bölümünde ise üç boyutlu olumsallık tablolarında i) değişkenlerden birinde kayıp olduğu ve ii) her iki değişkende de kayıp olduğu durumlar değerlendirilmiş ve kayıp gözlem varlığında kullanılan logaritmik doğrusal modeller uygulanmıştır. Modellere ait parametreler tahmin edilmiş ve bilgi kriterleri yardımıyla en iyi modellere karar verilmiştir. Veri yapısına en uygun olduğuna karar verilen model altında beklenen sıklıklar hesaplanmıştır.

Kalp Krizi x Hipertansiyon şeklinde oluşturulan iki boyutlu olumsallık tablosunda yalnızca hipertansiyon değişkeninde kayıp olması ve hem hipertansiyon hem de kalp krizi değişkenlerinde kayıp olması durumları incelenmiştir. Yalnızca hipertansiyon değişkeninde kayıp olduğu durumda uyum bulunan modele göre hipertansiyon değişkenindeki kayıp durumu tamamen bağımsızdır (MCAR). MAR ve NMAR türü kayıp gözlem olması durumunda modelin parametre sayısı göze sayısına eşit olduğu, dolayısıyla serbestlik derecesi sıfır bulunduğu için modele uyumu incelemek mümkün olmamaktadır. Bu nedenle üç model arasından hangisinin daha uygun olduğu ancak araştırmacının uzmanlığı sayesinde kayıp gözlem türüne karar verebilmesi ile mümkündür.

Hipertansiyon ve kalp krizi değişkenlerinin her ikisinde de kayıp olması durumunda elde edilen en iyi modele göre ise kalp krizi değişkenindeki kayıp durumu hipertansiyon değişkenine bağlıyken (MAR), hipertansiyon değişkenindeki kayıp durumu tamamen bağımsızdır (MCAR). İki örnekte de hipertansiyon değişkenindeki kayıp durumunun tamamen bağımsız olduğu modeller en iyi model olarak ortaya çıkmıştır ve özet tablolardan elde edilen odds oranları aynı sonucu vermiştir.

Hipertansiyon x Cinsiyet x Kalp Krizi şeklinde oluşturulan üç boyutlu olumsallık tablosunda yalnızca hipertansiyon değişkeninde kayıp olması ve hem hipertansiyon hem de cinsiyet değişkenlerinde kayıp olması durumları incelenmiştir. Yalnızca hipertansiyon değişkeninde kayıp olması durumunda elde edilen en iyi modele göre hipertansiyon değişkenindeki kayıp durumu kalp krizi değişkenine bağlıdır (MAR). Hipertansiyon ve cinsiyet değişkenlerinde kayıp olması durumunda elde edilen en iyi modele göre ise hipertansiyon değişkenindeki kayıp durumu hipertansiyon değişkeninin kendisine bağlıyken (NMAR), cinsiyet değişkenindeki kayıp durumu değişkenlerin tamamından bağımsızdır (MCAR). Yalnızca hipertansiyon değişkeninde kayıp olması ve hem hipertansiyon hem de cinsiyet değişkeninde kayıp olması durumlarında elde edilen odds oranları aynı sonucu vermiştir.

Uygulamanın üçüncü bölümünde, kayıp gözlem oranının model sonuçları üzerinde etkisini araştırmak amacıyla Krizi x Hipertansiyon tablosunun gözlenen sıklıkları sabit tutularak, hipertansiyondaki kayıp gözlem oranı %5, %10, %15 ve %20 olacak şekilde hipotetik tablolar oluşturulmuştur. Bu tablolara tek değişkende kayıp gözlem olduğu durumda kullanılan üç model uygulanmıştır. Kayıp gözlem oranının %5 ve %10 olduğu durumlarda hipertansiyondaki kayıp gözlemlerin kalp krizine bağlı olduğu ($b_{..}$) modeline uyum bulunmuştur. Kayıp gözlem oranının %15 ve %20 olduğu durumlarda ise modele uyum bulunamamıştır. Farklı kayıp gözlem oranlarında ($b_{..}$) modeline uyum sonuçları Çizelge 5.1'de özetlenmiştir. Kayıp gözlem oranı artıkça modele uyumun azaldığı, %15'den fazla kayıp gözlem olduğunda modele uyumun ortadan kalktığı söylenebilir.

Çizelge 5. 1. Farklı kayıp gözlem oranlarında ($b_{..}$) modeline uyum karşılaştırması

Kayıp Gözlem Oranı	G^2	p
%5	1,103	0,293
%10	2,896	0,089
%15	4,688	0,030
%20	6,915	0,009

Modele uyumları değişse bile model altında hesaplanan odds oranları tüm durumlarda gözlenen sıklıklar ile hesaplanan odds oranı ile aynı sonucu verir. Bunun sebebi kayıp gözlem türünün tamamen rastgele olarak ortaya çıkan ve ihmal edilebilir kayıp gözlem (MCAR) olmasıdır.

Bu tez çalışmasında tek değişkende ya da iki değişkende kayıp olması durumunda kullanılabilecek logaritmik doğrusal modeller incelenmiştir. Üç boyutlu olumsallık tablolarında tüm değişkenlerde kayıp gözlem bulunan durumda da kullanılabilen model yapıları mevcuttur. Üç değişkende de kayıp olduğu durumda, değişkenlerdeki kayıp gözlemlerin olası çeşitlerinin farklı kombinasyonları dikkate alınarak 64 farklı model oluşturulabilir. Bu modellerdeki karmaşık yapı hesaplama zorluklarını beraberinde getirmektedir.

Tez çalışması kapsamında araştırılan ve uygulama üzerinden tartışılan yöntemler kayıp gözlem içeren kategorik verilerle çalışan araştırmacılara yardımcı olacaktır. Logaritmik doğrusal model yaklaşımı sayesinde parametre tahminleri yapılarak beklenen sıklıklar elde edilebilir. Buradan elde edilen beklenen sıklıklar daha ileri analizlerde kullanılabilir.

Tek ya da iki değişkende kayıp olması durumunda kullanılan modeller kayıp gözlem çeşitlerini temel alır. Eğer araştırmacı kayıp gözlemin çeşidini kendisi saptayabiliyorsa tüm modelleri uygulamak yerine tablo yapısına ve kayıp gözlem çeşidine uygun olan modeli doğrudan uygulayabilir. Kendisi saptayamadığı durumda ise bu çalışmada olduğu gibi tablo yapısına uygun olan modelleri uygulayarak uyum bulduğu modeller içerisinden en uygun olanına bilgi kriterleri yardımı ile karar verebilir. En uygun olarak bulunan modeldeki yapı kayıp gözlemin türü hakkında da araştırmacıya bilgi verecektir.

Bu çalışmanın devamı olarak sınır çözümleri yöntemleri ve üç boyutlu tablolarda değişkenlerinin tamamında kayıp gözlem olması durumunda kullanılan logaritmik doğrusal modellere ilişkin geniş kapsamlı bir çalışma yapılabilir. Kayıp gözlem içeren farklı yapılarda tabloların üretildiği ve modellere uyumun incelendiği geniş çapta bir simülasyon çalışması yapılabilir.

KAYNAKÇA

Afifi, A. A. and Elashoff, R. M., Missing Observations in Multivariate Statistics: I. Review of the Literature, *JASA* 61, 315 (**1966**): 595-604.

Agresti, A., *Categorical Data Analysis.* John Wiley & Sons, **2002**.

Aktaş Altunay, S., Yılmaz, A. E., Bahçecitapar, M., Karabenli, L. B. *SPSS ve R Uygulamalı Kategorik Veri Çözümlemesi,* Ankara: Seçkin Yayıncılık, **2021**.

Allison, P. D., Missing Data, *Series: Quantitative Applications in the Social Sciences*, 72-89, SAGE University Paper, **2001**.

Alpar, R. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*, 131-164. Ankara: Detay Yayıncılık, **2021**.

Baker, S. G. and Laird, N. M., Regression Analysis for Categorical Variables with Outcome Subject to Nonignorable Nonresponse, *Journal of The American Statistical Association* 83, 401 (**1988**): 62-69.

Baker, S. G., Rosenberger, W. F. and Dersimonian, R., Closed-form Estimates for Missing Counts in Two-way Contingency Tables, *Statistics in Medicine, 11, 5,* (**1992**): 643-567.

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

Celuex, D., Diebolt, J., The SEM Algorithm: A Probabilistic Teacher Algorithm Derived from The EM Algorithm for The Mixture Problems, *Computational Statistics* 2 (**1985**): 73-82.

Chen, T. and Fienberg, S. E., The Analysis of Contingency Tables with Incompletely Classified Data, *Biometrics* 32, 1 (**1976**): 133-144.

Çağılcı, T. and Danacıoğlu, N., Log-lineer Modeller ve Kadına Yönelik Şiddet Üzerine Bir Uygulama, *Sinop Üniversitesi Fen Bilimleri Dergisi* 5, 1 (**2020**): 1-12.

Dayanıklı, A.S., *Kayıp Veri Analizi*, https://ravenfo.com/2021/06/06/kayip-veri-analizi/ (Erişim tarihi: **Mayıs 2023**).

Dempster, A. P., Laird, N.M. and Rubin, D. B., Maximum Likelihood from Incomplete Data via The EM Algorithm, *Journal of The Royal Statistical Society: Series B (Methodological)* 39, 1 (**1977**): 1-22.

Freedman, D. A. and Wachter, K. W., Methods for Census 2000 and Statistical Adjustments, *The SAGE Handbook of Social Science Methodology. London: Sage*, **2007**: 232-245.

Fuchs, C., Maximum Likelihood Estimation and Model Selection in Contingency Tables with Missing Data, *Journal of the American Statistical Association* 77, 378 (**1982**): 270-278.

Gad, A. M. and Ahmed, A. S., Analysis of Longitudinal Data with Intermittent Missing Values Using The Stochastic EM Algorithm, *Computational Statistics & Data Analysis* 50, 10 (**2006**): 2702-2714.

Ghosh, S. and Vellaisamy, P., Closed-form Estimates for Missing Counts in Multidimensional Incomplete Tables, *Hacettepe Journal of Mathematics and Statistics* 53, 3 (**2024**): 803-822.

Ghosh, S. and Vellaisamy, P., On the Occurrence of Boundary Solutions in Two-way Incomplete Tables, *Revstat-Statistical Journal* 18, 1 (**2020**): 89-108.

Ghosh, S. and Vellaisamy, P., Evaluation of Missing Data Mechanisms in Two and Three Dimensional Incomplete Tables, *Journal of the Korean Statistical Society* 48, 2 (**2019**): 297-313.

Ghosh, S. and Vellaisamy, P., On The Occurrence of Boundary Solutions in Multidimensional Incomplete Tables, *Statistics & Probability Letters* 119 (**2016**): 63-75.

Golovenkin, S. E., Shulman, V. A., Rossiev, D.A., Sherstenya, P.A., Nikulina, S. Yu., Orlova and Yu. V., Voino Yasenetsky, V.F., Myocardial Infarction Complications, *UCI Machine Learning Repository*, **2020**. https://archive.ics.uci.edu/dataset/579/myocardial+infarction+complications (Erişim tarihi: **Eylül 2023**).

Hartley, H. O., Maximum Likelihood Estimation from Incomplete Data, *Biometrics* 14, 2 (**1958**): 174-194.

Heijden, V.d., De Falguerolles P.G. and Leeuw, J.De., A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Loglinear Analysis, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 38, 2 (**1989**): 249-273.

Hooda, D. S. and Barak, M. S., Estimation of Missing Data in Design of Experiment and Contingency Table, *SN Applied Sciences* 1, 7 (**2019**): 670.

Hooda, D. S. and Barak, M.S., Estimation of Missing Values in Fuzzy Matrices (FM) and Interval-valued Fuzzy Matrices (IVFM), *Life Cycle Reliability and Safety Engineering* 9 (**2020**): 241-245.

Howell, D. C., The Treatment of Missing Data, *The Sage Handbook of Social Science Methodology*, **2007**: 208-224.

Kale, F., Kayıp Veri ile Baş Etme Yöntemleri, *Ordu Üniversitesi Fen Bilimleri Enstitüsü*, **2020**.

Kim, S., Jeon, S. and Kim, D., On Log-linear Modeling for an Incomplete Two-way Contingency Table with One Variable Subject to Nonresponse, *Communications in Statistics-Simulation and Computation* 49, 4 (**2020**): 973-988.

Kim, S., Park, Y. and Kim, D., On Missing-at-random Mechanism in Two-way Incomplete Contingency Tables, *Statistics & Probability* 96 (**2015**): 196-203.

Köse, İ. A. and Öztemur, B., Kayıp Veri Ele Alma Yöntemlerinin t-Testi ve ANOVA Parametreleri Üzerine Etkisinin İncelenmesi, *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, **2014**: 400-412.

Little, R. J. A. and Rubin, D., *Statistical Analysis with Missing Data*, John Wiley & Sons, **1987**.

Molenberghs, G., Beunckens, C., Sotto, C. and Kenward, M. G., Every Missingness Not at Random Model Has a Missingness at Random Counterpart with Equal Fit, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 70, 2 (**2008**): 371-388.

Öztürk, E., Zengin, Y., Kaşıkçı, M. and Coşgun, E., The Effect of Missing Data Mechanisms on Deep Learning in Binary Classification: A Simulation Study, *Türkiye Klinikleri Biyoistatistik* 15, 1 (**2023**): 1-12.

Park, T., An Approach to Categorical Data with Nonignorable Nonresponse, *Biometrics*, **1998**: 1579-1590.

Park, T. and Brown, M. B., Models for Categorical Data with Nonignorable Nonresponse, *Journal of The American Statistical Association* 89, 425 (**1994**): 44-52.

Park, Y., and Choi, B.S., Bayesian Analysis for Incomplete Multi-way Contingency Tables with Nonignorable Nonresponse, *Journal of Applied Statistics* 37, 9 (**2010**): 1439-1453.

Pembegül, A., İhmal Edilemeyen Kayıp Veri Varlığında Olumsallık Tablo Çözümlemeleri, *Hacettepe Üniversitesi Fen Bilimleri Enstitüsü*, **2009**.

Peng, C. Y. J., Harwell, M., Liou, S. M. and Ehman, L. H., Advances in Missing Data Methods and Implications for Educational Research, *Real Data Analysis*, **2006**: 3178, 102.

Perlman, M. D. and Wu, L., Lattice Conditional Independence Models for Contingency Tables with Non-monotone Missing Data Patterns, *Journal of Statistical Planning and Inference* 79, 2 (**1999**): 259-287.

Phillips, M. J., Contingency Tables with Missing Data, *Journal of The Royal Statistical Society: Series D (The Statistician)* 42, 1 (**1993**): 9-18.

Rochani, H. D.,Vogel, R. L., Samawi, H. M. and Linder, D. F., Estimates for Cell Counts and Common Odds Ratio in Three-way Contingency Tables by Homogeneous Log-linear Models with Missing Data, *AStA Advances in Statistical Analysis* 101, no. 1 (**2017**): 51-65.

Rubin, D. B., Inference And Missing Data, *Biometrika* 63, 3 (**1976**): 581-592.

Rubin, D. B., Multiple Imputation for Nonresponse in Surveys, *John Wiley & Sons*, **1987**.

Schafer, J. L., *Analysis of Incomplete Multivariate Data*, CRC press, **1997**.

Schafer, J. L., and Olsen, M. K., Multiple Imputation for Multivariate Missing-data Problems: A Data Analyst's Perspective, *Multivariate Behavioral Research* 33, 4 (**1998**): 545-571.

Wu, L., Reeves, J. and Perlman, M. D., Some Model-restricted Shrinkage Estimators for Contingency Tables with Missing Data, *Metrika* 50 (**2000**): 221-245.

Yazıcı, F., EM Algortiması ve Uzantıları, *Hacettepe Üniversitesi Fen Bilimleri Enstitüsü*, **2005**.