

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

**NORMAL, MULTİNOMİAL, ÜSSEL (EXPONENTIAL) VE GAMMA
DAĞILIM GÖSTEREN VERİ YAPILARI VE EKSİK VERİ TİPLERİNDE
(MAR, MCAR, MNAR) TAMAMLAMA ALGORİTMALARININ PARAMETRE
TAHMİNLERİ ÜZERİNE ETKİLERİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Serdar İNAN
DANIŞMAN: Dr. Öğr. Üyesi Şakir İŞLEYEN

VAN-2019

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

**NORMAL, MULTİNOMİAL, ÜSSEL (EXPONENTIAL) VE GAMMA
DAĞILIM GÖSTEREN VERİ YAPILARI VE EKSİK VERİ TİPLERİNDE
(MAR, MCAR, MNAR) TAMAMLAMA ALGORİTMALARININ PARAMETRE
TAHMİNLERİ ÜZERİNE ETKİLERİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Serdar İNAN

VAN-2019

KABUL VE ONAY SAYFASI

İstatistik Anabilim Dalı'nda Dr. Öğr. Üyesi Şakir İŞLEYEN danışmanlığında, Serdar İNAN tarafından sunulan Normal, Multinomial, Üssel (Exponential) ve Gamma Dağılım Gösteren Veri Yapıları ve Eksik Veri Tiplerinde (MAR, MCAR, MNAR) Tamamlama Algoritmalarının Parametre Tahminleri Üzerine Etkileri” isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince/...../..... tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile başarılı bulunmuş ve tezi olarak kabul edilmiştir.

Başkan:.....

İmza:

Üye:.....

İmza:

Üye:.....

İmza:

Üye:.....

İmza:

Üye:.....

İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve sayılı kararı ile onaylanmıştır.

İmza

.....
Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

(İmza)

Serdar İNAN



ÖZET

NORMAL, MULTİNOMİAL, ÜSSEL (EXPONENTIAL) VE GAMMA DAĞILIM GÖSTEREN VERİ YAPILARI VE EKSİK VERİ TİPLERİNDE (MAR, MCAR, MNAR) TAMAMLAMA ALGORİTMALARININ PARAMETRE TAHMİNLERİ ÜZERİNE ETKİLERİ

İNAN, Serdar

Yüksek Lisans Tezi, İstatistik Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Şakir İŞLEYEN

Haziran 2019, 67 sayfa

Eksik veri problemi araştırmacıların istatistiksel analiz yaparken sürekli karşılaştıkları sorunlardan biridir. Eksik veri, çalışmayı yürüten bireyden, verisi toplanan kişilerin bilerek cevap vermemesinden, verilerin gözlenmemesinden, veriyi kaydetmek için kullanılan ekipmanlardan ve bunun gibi nedenlerden kaynaklanabilmektedir. Verinin eksik olması en büyük sorunlardan biridir. Zira hemen hemen tüm istatistiksel analizler eksiksiz bir veri setine ihtiyaç duymaktadır.

Çalışmada dört dağılım türüne göre veriler üretilmiş, bu veriler eksik veri mekanizmaları olan MCAR, MAR ve MNAR tiplerine göre belli oranlarda eksiltilmiş ve bu eksiltelen veriler eksik veri analiz yöntemlerine göre tamamlanmışlardır. Türetilen veri, eksiltelen veri ve tamamlanan verilerin parametreleri hesaplanmış ve çıkan bu değerlerin karşılaştırmaları yapılmıştır. Sonuçta çoklu atama yönteminin diğer yöntemlere göre daha etkin ve tarafsız parametre tahminleri yaptığı tespit edilmiştir.

Anahtar kelimeler: Çoklu atama, Eksik veri, Eksik veri mekanizmaları, Parametre tahmini.

ABSTRACT

THE EFFECTS OF COMPLETION ALGORITHMS ON PARAMETER ESTIMATES IN NORMAL, MULTINOMIAL, EXPONENTIAL AND GAMMA DISTRIBUTED DATA STRUCTURES AND MISSING DATA TYPES (MCAR, MAR, MNAR)

İNAN, Serdar

M.Sc. Thesis, Department of Statistics

Supervisor: Asst. Prof. Dr. Şakir İŞLEYEN

June 2019, 67 pages

The missing data problem is one of the main problems that researchers constantly encounter when performing statistical analysis. Missing data may result from the individual conducting the work, the inability of the data to be collected, the inability to observe the data, the equipment used to record the data, and the like. The biggest problem with the lack of data is the need for a complete set of data for almost all statistical analyses.

In this study, data were produced according to four types of distribution, these data were decreased in certain ratios according to MCAR, MAR and MNAR types which are incomplete data mechanisms and these minimized data were completed according to missing data analysis methods. Calculated data, deducted data and parameters of the completed data were calculated and compared with these values. As a result, it has been determined that multiple imputation method makes more effective and neutral parameter estimation compared to other methods.

Keywords: Missing data, Missing data mechanisms, Multiple imputation, Parameter estimates.

ÖN SÖZ

Bu tezin hazırlanmasında daima bilgi ve tecrübesi ile bana destek veren tez danışmanım Sayın Dr. Öğr. Üyesi Şakir İŞLEYEN'e teşekkür ederim. Ayrıca Prof. Dr. H. Eray ÇELİK ve Doç.Dr. Hamit MURTAGİOĞLU hocalarıma, her türlü desteği bana veren Prof.Dr. Kazım ÖZDAMAR, Prof. Dr. Fevzi ERDOĞAN, Dr. Öğr. Üyesi Yıldırım DEMİR, Dr. Öğr. Üyesi Sedat ÇELİK hocalarıma, her daim yanı başımda olan Öğr. Gör. Necat ERDOĞAN ve Öğr.Gör. Adem ŞEHİTOĞLU dostlarıma, beni bu yola teşvik eden, her zaman her türlü desteği veren büyüğüm Dr. Öğr. Üyesi Abdullah OĞRAK ağabeyime ve aileme teşekkürlerimi sunarım.

2019

Serdar İNAN



İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	iii
ÖN SÖZ.....	v
İÇİNDEKİLER.....	vii
ÇİZELGELER LİSTESİ	ix
ŞEKİLLER LİSTESİ	xi
SİMGELER VE KISALTMALAR	xiii
1. GİRİŞ	1
2. GENEL BİLGİLER	5
2.1. Veri Matrisi.....	5
2.2.Eksik Veri Tanımı	6
2.3. Eksik Verinin Tarihsel Gelişimi	8
2.4. Eksik Veri Mekanizmaları (Türleri).....	9
2.4.1. Tamamen rastgele eksik veri (missing completely at random-MCAR).....	10
2.4.2. Rastgele eksik veri (missing at random -MAR).....	11
2.4.3. Rastgele olmayan eksik veri (missing not at random - MNAR).....	13
3. EKSİK VERİLERİ YÖNETME YÖNTEMLERİ.....	19
3.1. Geleneksel Yöntemler	21
3.1.1. Liste boyunca silme (listwise deletion-dizin silme)	21
3.1.2. Çift silme (pairwise deletion-ikili silme)	24
3.1.3. Ortalama atama (mean imputation-ortalama değerlendirme)	27
3.1.4. Regresyon atama (regression imputation).....	29
3.1.5. Tahmini ortalama eşleştirme (sıcak deste- hot deck imputation).....	32
3.1.6. Soğuk deste(cold desk imputation).....	34
3.1.7. En yakın komşu yöntemi(nearest neighbour).....	34
3.1.8. Son gözlemi ileri taşıma(last observation carried forward).....	34
3.2. Modern Atama Yöntemleri.....	35
3.2.1. Çoklu atama (multiple imputation)	35

3.2.2. Beklenti maksimizasyonu (em: expectation maximization).....	37
4. MATERYAL ve YÖNTEM	39
4.1. Veri Üretimi.....	39
4.2. Veri Eksiltilmesi	39
4.4. Veri Tamamlanması.....	40
4.4. Parametrelerin Hesaplanması	40
5. BULGULAR.....	41
5.1. Normal Dağılım.....	41
5.2. Üstel Dağılım	47
5.3. Gamma Dağılımı	53
5.4. Multinomial Dağılım	58
6. TARTIŞMA ve SONUÇ	59
KAYNAKLAR.....	63
ÖZ GEÇMİŞ.....	67

ÇİZELGELER LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Zekâ puanı ile MCAR, MAR ve MAR mekanizmaları oluşturulmuş iş performans puanları.....	15
Çizelge 2.2. MCAR ve MAR mekanizmaları arasındaki farklar	17
Çizelge 3.1. Liste silme için örnek veri tabanı	24
Çizelge 3.2. Liste silme işlemi sonrası örnek veri tabanı.....	25
Çizelge 3.3. Eşlerin silinmesi öncesi veri tabanı	28
Çizelge 3.4. Eşlerin silinmesi sonrası veri tabanı	28
Çizelge 3.5. Ortalama atama öncesi veri tabanı	31
Çizelge 3.6. Ortalama atama sonrası veri tabanı	31
Çizelge 3.7. Regresyon atama öncesi veri tabanı	33
Çizelge 3.8. Regresyon atama öncesi veri tabanı	34
Çizelge 5.1. Normal dağılımda MCAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	45
Çizelge 5.2. Normal dağılımda MAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	47
Çizelge 5.3. Normal dağılımda MNAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	49
Çizelge 5.4. Üstel dağılımda MCAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	51
Çizelge 5.5. Üstel dağılımda MAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	53
Çizelge 5.6. Üstel dağılımda MNAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	54
Çizelge 5.7. Gamma dağılımında MCAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	56

Çizelge**Sayfa**

Çizelge 5.8. Gamma dağılımında MAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	58
Çizelge 5.9. Gamma dağılımında MNAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri	59



ŞEKİLLER LİSTESİ

Şekil	Sayfa
Şekil 2.1. Veri matrisi	5
Şekil 2.2. MCAR mekanizması ilişki gösterimi	11
Şekil 2.3. MAR mekanizması ilişki gösterimi.....	13
Şekil 2.4. MNAR mekanizması ilişki gösterimi.....	15
Şekil 3.1. Kilo ve boy regresyon grafiği	34
Şekil 3.2. Çoklu atama aşamaları	39



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklama
A.O	Aritmetik ortalama
Kor.	Korelasyon (ilişki) katsayısı
Kov.	Kovaryans
MAR	Missing At Random-Rastgele Eksik Veri
MCAR	Missing Completely At Random- Tamamen Rastgele Eksik Veri
MNAR	Missing Not At Random- Rastgele Eksik Olmayan Veri
S.S.	Standart Sapma
R²	Çoklu Belirlilik Katsayısı (Multiple R ²)

1. GİRİŞ

Bilimsel alanda yaşanan gelişmeler asırlardır sürmektedir. İnsanlığın yaşam mücadelesi, sürekli gelişerek yaşamı daha kolay hale getirmenin anahtarı olmuştur. Geçmişte yaşamı kolaylaştırma ya da meraklar sonucunda ortaya çıkan keşifler ve elde edilen bilgiler, zamanla kendi yollarını bulmuş, akıl ve düşünce ile birleşmiştir. Bu, bize üzerinde yaşadığımız Dünya'yı ve imkânlarını daha yakından tanıma, daha fazla bilgi sahibi olma fırsatını vermiştir. Bilgiye erişme ve sahip olma arzusu, insanoğlunun vazgeçilmez isteklerindedir. Bu nedenle bilgi günümüzün sahip olunabilecek en önemli unsurdur. Bilgiye ulaşmak için yapılan çalışmalarda araştırmacıların başlangıç noktaları, geçmişteki deneyimlerdir. Kaynaksız bir kitap olmayacağı gibi yeni bir şey için düşünsel ya da deneysel bir birikimin olmaması da mantıksızdır. Yeni bir şeyi ortaya çıkartmak için eskiden yararlanmak en geçerli yöntemlerden biridir. Geleceği ve olacağı kestirebilmek adına faydalanılan istatistik, geçmiş birikimlerle gelecek için çalışmalar yapmak amacıyla kullanılan, kendine özgü yöntemleri sayesinde hemen hemen her bilim dalının uygulamalarında karşılaşılabilen bir disiplindir (Toka, 2012; Nartgün, 2015) .

İstatistik, geçmişteki birikimler yoluyla toplanmış olan tüm bilgiyi çözümleyerek daha sonraki süreçte alınacak kararlara yardımcı olacak sonuçları verir. Sürecin gerçekleşme ihtimali ile ilgili bilgiler sağlar. Aynı zamanda deneysel düzenlemelerle elde edilen sonuçların geçerliliğini ve doğruluğunu araştırır. Yeni yöntemler oluşturularak bilim dallarının birçok alanına uygulanabilir ve uygun modeller tanımlanabilir. İstatistiğin gerek diğer bilim dallarıyla olan ilişkisinden gerekse sonuç değerlendirmelerinden dolayı önemli olduğu görülmektedir. Dolayısıyla yapılan her inceleme sonucunda elde edilecek bilgiye güvenebilmek için sağlıklı ve güvenilir veriyle çalışmak, veriye uygun istatistiksel çalışmalar yapmak gerekmektedir (Toka, 2012) .

Elde edilen bir veri kümesinde geçmiş bilgilerin kaybı, yok olmaları ve bazı imkânsızlıklardan elde edilememesi büyük bir problemdir. Aynı şekilde deneysel düzenlemelerde kayıt altına alınmayan, maddi gerekçelerle tekrarlanamayan bilginin eksikliği de büyük bir problem oluşturmaktadır. Bu noktadan itibaren istatistiksel yöntemlerdeki bilgi eksikliği göz önüne alınarak eksik değerleri kestirebilmek için çeşitli çalışmalar üzerinde durulmuş ve bu süreç sonunda eksik veri durumunda istatistiksel

yöntemlerin nasıl ele alınması gerektiğine dair sorulara cevap verilmeye başlanmıştır (Demir, 2012) .

Eksik veriler, insan tarafından toplanan verilere dayanan hemen hemen tüm araştırma alanlarında yaygın bir sorundur. Verilerin eksik olması, genellikle bir veri setindeki bazı vakaların eksik veriye sahip olduğu, yani bazı durumlarda veri setinde yer alan bazı değişkenlerin tümünün, bazı durumlarda ise tümünün değil de belirli bir kısım değişkenin bilgileri eksik olduğunda ortaya çıkan durum anlamına gelir. Bilginin eksik olmasının birkaç nedeni olabilir; bazı katılımcılar bir ankette yer alan bazı soruları cevaplamayı reddetmiş veya unutmuş olabilir, bilginin toplandığı yer belirli bir değişken için veri toplama kapasitesine sahip olmayabilir veya bir veri setindeki mevcut verilerin gerçek olmadığı veya hatalı olduğu tespit edilmiş olabilir. Eksik verinin, yanıt vermeme durumundan farklı olduğunu vurgulamak önemlidir, çünkü eksik veri her durumda en azından bazı değişkenlerin gözlenmesini gerektirir. Bu nedenle, eksik verilerin yol açtığı hatalar, temel olarak, cevaplanmayan önyargı nedeniyle oluşabilecek hatalardan farklıdır. Eksik verili veri setleri problemlidir, çünkü veri analizi için yaygın olarak kullanılan istatistiksel yöntemlerin çoğunun çalışması, verilerin eksiksiz olması veri setinin herhangi bir eksik veriye sahip olmamasını ve düzgün çalışması ilkesine dayanır. Bu nedenle, eksik verilerin varlığını göz ardı etmek, bu verileri kullanarak yapılan analizlerin sonuçlarını etkileyecek ve taraflı veya yanlış sonuçlara yol açabilecektir (Enders, 2010; Toka, 2012).

Veri kümelerinde var olan eksik verilerle baş edebilmek için eksik veriyi eksik verili kaydı silmek, göz ardı etmek, sık geçen ifade ile doldurmak veya ilgili kayıt ya da niteliğin satır, sütun ortalaması ile doldurulması gibi basit klasik yöntemler karmaşık hesaplama yöntemlerinin yerine kullanılmaktadır (Cheng ve Victoria, 2002; Meng ve Shi, 2012). Ancak bu gibi basit klasik yöntemler eksik verileri yok sayarak verimi düşürmekte bununla beraber var olan veriyi yanlışlaştırarak sistematik anlamda kalitesizleştirmektedir (Little ve Rubin, 1987; Pelckmans ve ark. 2005; Nelwamondo ve Marwala, 2008). Eğer eksik veriler düzgün bir şekilde ele alınıp hesaplanmazsa araştırma sonuçlarının geçerliliğinin düşebileceği gibi bazı durumlarda araştırma başarısızlığa bile uğrayabilmektedir (Genc ve ark., 2010; Di Nuovo, 2011; Bergmeir ve Benitez, 2012).

Eksik verilerin ele alınması için hali hazırda R Language’da Mice (Van Buuren ve Groothuis-Oudshoorn 2011), Amelia II (Honaker ve ark. 2011), ve NORM R (Novo

ve Schafer 2013) ve Proc MI for SAS (SAS Institute 2015) gibi çok sayıda farklı istatistiksel istatistiksel araç ve paket programları bulunmaktadır. Bu araçlar ve paket programlar farklı türden töhmet yöntemlere, yani eksik verileri yeni değerlerle değiştiren yöntemlere dayanır, böylece analizlerde tam veri istatistik yöntemlerinin kullanılmasına izin verir. Bununla birlikte, bu mevcut paketler ve araçlar, bazı zayıf yönlere sahiptir. Başlıca zayıf noktalardan biri, bu mevcut paketlerin ve araçların geniş kapsamlı istatistiksel bilgiler olmadan anlaşılması zor ve kullanımı zor olmasıdır. Ek olarak, bunların çoğunun hesaplama açısından yavaş olması ve/veya kategorik ve sıra değişkenlerin her zaman makul olmayan bazı varsayımlar olmadan ele alınmasına izin vermemesidir (Allison, 2009).

Eksik veriler, veri toplamının doğal bir parçası olarak kabul edilen bir olgudur, çünkü veri setindeki bazı vakalarla ilgili belirli bilgilerin elde edilememesi olağandır. Bununla birlikte, bir veri setindeki diğer tüm değerler gibi, eksik veriler, belirli bir verinin gözlenip gözlenmeyeceğini belirleyen bir tür eksik veri mekanizmasından kaynaklanmış olabilir. Bu eksik veri mekanizmasının işleyiş süreci daha sonra, eksik verilerin ele alınması için farklı yöntemlerin etkinliğini etkileyecektir (King ve ark, 2001). Bu yüzden, eksik verileri ele almada farklı yöntemleri tartışmak ve değerlendirmek için, bu farklı eksik veri mekanizmalarının nasıl işleyebileceği ile ilgili bazı genel varsayımların yanı sıra, eksik verileri ele almak için farklı yöntemler üzerindeki etkileri tartışılmalıdır (Allison, 2012).

Çalışmamızda değişik araştırma alanlarında sıklıkla rastlanan eksik veri durumlarında veri kaybının önlenmesi bakımından önem taşıyan eksik verilerin tamamlanması ve geçerli, güvenilir parametre tahminlerine ulaşılabilmesi bakımından bilim ve uygulama alanlarında büyük önem taşıyan bilgiler ortaya koyabilmek amaçlanmıştır. Bu kapsamda ele alınabilecek eksik veri tamamlama yöntemleri incelenerek hangisinin istatistikleri parametrelere daha çok yaklaştıracığı ve hangisinin veri israfına yol açmayacağıın tespit edilmesi amaçlanmıştır.



2. GENEL BİLGİLER

Bu bölümde ilk olarak eksik veri ve yapısı hakkında genel bilgiler verilecektir. Daha sonra ise eksik veri mekanizmalarından bahsedilecektir.

2.1. Veri Matrisi (Data Matrix)

İstatistik analizlerde esas itibariyle standart istatistiksel yöntemlerin doğasından kaynaklanan eksik veri sorunu, önemli bir tartışma alanıdır. Standart istatistiksel yöntemler, dikdörtgensel veri setlerinin analizi için geliştirilmiştir. Bir matris şeklinde hazırlanan bu veri setlerinde satırlar gözlemleri, sütunlar ise değişkenleri temsil etmektedir. Bir değişkene yönelik bir gözlemin bulunmaması durumunda, bu gözlemi temsil eden hücre boş kalır ve eksik veri oluşur. Sonuçta tipik bir veri setinin eksik veri içermesi, veri setinde yer alan bazı değişkenlere ilişkin bilgilerin ya da gözlemlerin bulunmaması anlamına gelmektedir (Little ve Rubin, 1987).

İstatistik analizlerde veriyi daha kolay gösterebilmek için veri matrisi kullanılır. Veri matrisi, gözlemin bulunduğu satırlar ile değişkenin bulunduğu sütunlardan oluşan değerler (bilgiler) topluluğu olarak kabul edilir ve en genel haliyle Şekil 2.1. 'deki gibidir:

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Şekil 2.1. Veri matrisi.

Satırlardaki her gözlem, sütunlardaki özellikleri ifade eden sayısal değerle temsil edilerek veri kümesini oluşturur. Veri matrisleri genellikle alfabenin harfleri ile temsil edilirken matrisin herhangi bir gözlemi i indisiyle, herhangi bir değişkeni ise j indisiyle gösterilir. Bu gösterimler matematiksel anlatımlarda ve eşitliklerde kolaylık sağlamak amacıyla kullanılmaktadır. Gösterimlerde X matrisinin x_{ij} . elemanı, i . gözlemin j .

değişken için aldığı değeri ifade etmektedir. X matrisinin diğer bir gösterim şekli ise (x_{ij}) 'dir.

2.2. Eksik Veri Tanımı

Little ve Rubin (2002), veri matrisinin dikdörtgen biçimli bir görünüme sahip olmasından dolayı, standart istatistiksel yöntemlerin dikdörtgen veri biçimini çözebilecek şekilde geliştirildiğini açıklamışlardır. Veri matrisi bir dikdörtgene benzeyeceğinden dolayı herhangi bir eksiklik olması durumunda veri yapısının dikdörtgen biçiminin bozulacağını ve veri matrisindeki bazı değerlerin eksilmesinin eksik veri sorununu ortaya çıkaracağını belirtmişlerdir.

Longford (2006), eksik veri için en basit tanımın, planlanan veri kümesi ile elde edilen veri kümesi arasındaki farklılık olduğunu bildirmiştir. Bir gözlemin tümünün ya da herhangi bir değişkenin tüm değerlerinin elde edilememesinin eksik veri olarak isimlendirilemeyeceğini belirten bazı araştırmacılar (özellikle sosyal bilimler ve eğitim bilimleri), bu tanımın çok kapsayıcı olduğunu belirtmişleridir.

Allison (2000), veri matrisindeki bazı gözlemlerin birkaç değişkenine karşılık gelen değerlerin elde edilememiş olmasını eksik veri olarak tanımlamıştır. Allison, bütün gözlemler için bir değişkenin elde edilmemiş olmasını eksik veri olarak adlandırmamış, gizli değişken ya da gözlenmemiş değişken olarak adlandırmıştır. Aynı şekilde bazı gözlemler için bütün değişkenlerin eksik olması durumunu gözlemin elde edilememesi sorunu olarak açıklamıştır.

Bal (2003), araştırmalarda bulunan eksik verilerin, bilgi eksikliğini temsil edeceğini, dolayısıyla bunun bilgi eksikliğine neden olacağını dile getirmiştir. Aynı zamanda istatistiksel analizlerde kullanılan programlar tam veri setlerine göre düzenlendiğinden dolayı, toplanan verilerde bulunan eksiklerin istatistiksel analizlerde önemli sorunlara neden olacağını vurgulamıştır. Bu yüzden eksik verilerden kaynaklanan problemlerin yaşanmaması için araştırmacıların tam veri setleri ile çalışması gerektiğini belirtmiştir.

Elde bulunan bilgiden en yüksek seviyede yararlanarak çözümler elde eden istatistik bilimi için uygun bir çözümlenmeyi seçmek kadar tüm verinin anlattıklarını en sağlam bilgiye dönüştürmek önemlidir. Eksik veriye sahip veri matrisleri istatistiksel

çözümlemelerde kullanılan paket programlarında sorun çıkarttığından dolayı bazı gözlemler elde edilebilecek durumda olmasına rağmen sonuçlarda kullanılmama durumu ile karşılaşılmaktadır. Bu da analiz sonuçlarına bilgi kaybı olarak girer. Elde edilen veri setinde eksik gözlemler(eksik veriler) olduğunda, eldeki verilerin analizinde büyük zorluklar yaşandığından araştırmacılar eksik veri analizlerini geliştirmişlerdir. Eksik veri analizine eksik veri içeren herhangi bir veri seti için istatistik analiz uygulanmadan önce başvurulması gereken bir ön analiz olarak bakılmaktadır (Bal, 2003). Eksik veri analizi ile veri setindeki eksik veriler, uygun değerlerle tamamlanır ve tamamlanmış veri seti ile istatistiksel işlemler yapılır (Little ve Rubin,2002).

Araştırmalarda eksik veri sorununun ancak 1970 sonrasında tartışılmaya başlanması ve eksik veri yöntemlerinin ancak 1990 sonrasında istatistiksel yazılımlara yansması, geliştirilmiş birçok eksik veri yönteminin henüz beklentileri karşılayacak düzeyde olmamasını bir ölçüde açıklamaktadır.

Eğer eksik veri miktarı fazlaysa elde edilen sonuçların güvenilirliği, genellenebilirliği ve yapılacak analizler bundan önemli derecede etkilenecektir (McKnight., 2007).

Eksik veri analizleri için kullanılacak olan klasik ve modern istatistiksel yöntemlerin hemen hemen hepsi için önemli bir sorun oluşturur çünkü tüm yöntemler veri setinin eksiksiz olduğu varsayımı altında geliştirilmiştir. (Pigott, 2001; Allison, 2003; Osborne, 2013).

Allison (2002)'a göre, araştırmacılar genelde, eksik verilerin gözlenen ölçmelerde bir farklılığa yol açmadığını kanıtlama eğilimindedir. Örneğin gelirlerini beyan eden ve etmeyen katılımcılar arasında, incelenen değişkenler açısından anlamlı bir farklılık oluşmadığına yönelik kanıt sağlama çabası oldukça yaygındır (Demir, 2012). Rubin'e göre çoğu araştırmada eksik verilerin analiz dışı bırakılmasının temelinde bu verilerin, bir takım varsayımlara bağlı olarak 'ihmal edilebilir (ignorable)' görünmesi yatmaktadır. Eksik verilerin ihmal edilebilir olduğunun kanıtlanması, eksik değerlerle, rastgele olarak belirlenen gözlenen verilerin örtüşeceği varsayımına dayanır. Dolayısıyla tam ya da eksiksiz veri seti ile eksik veri içeren veri setinin dağılımları ve bu veriler üzerinden yapılacak çıkarımlar arasında kayda değer bir fark olmadığı şeklindeki görüş desteklenmiş olur (Demir, 2012).

2.3. Eksik Verinin Tarihsel Gelişimi

İstatistiksel yöntemlerin gelişmesi ile eksik veri sorunu ortaya çıkmış ve bu konuda çeşitli çalışmaların yapılmasına başlanmıştır. Eksik veriyi analiz çalışmaları, her ne kadar 1930'larda başlamış olsa da eksik verinin bir merak ve ilgi alanı olması Rubin (1976)'in eksik veri düzeneklerini sınıflandırmasıyla artmıştır. Little ve Rubin (1987)'in çalışması da eksik veri analizleri için önemli bir kaynak olmuştur. Bu nedenle, Rubin (1976)'in çalışması eksik veri için dönüm noktası olarak görülmektedir.

İlk defa eksik veri ilkeleri tanımını kullanan ve çok değişkenli normal dağılıma sahip veri kümesini tam veri kısmı ve tamamlanmamış veri kısmı olarak ayırıp elde edilecek parametre değerleri için eşitlikler oluşturan Orchard ve Woodbury (1972) eksik verinin elde edilmesi için olabilirlik fonksiyonu üzerinden çalışma yapmışlar ve en büyük olabileceği noktalardaki değerleri eksik gözlemlere yüklemişlerdir. Dempster ve ark. (1977), Orchard ve Woodbury (1972)'nin Beklenti Maksimizasyon (Expectation Maximization – EM) algoritmasının başlangıç teorisini ortaya attıklarını belirtmişlerdir. Little ve Rubin (2002), EM algoritmasından önceki çalışmaların sürekli matris tersi alma gerektirdiğini, EM algoritmasının bu gereksinime ihtiyaç duymadığından üstün olduğunu açıklamışlardır.

Rubin (1976), eksik veriyi üç durumda incelemiştir. Bu inceleme sonunda çözümlenmenin hangi durumda yapılabileceğini veya yapılamayacağını göstermiştir. Bu durumlar veri düzenekleri olarak isimlendirilmiştir.

Little ve Rubin (1987), eksik veri yapısının nasıl olabileceği konusunda önceki çalışmaları inceleyerek örnekler vermişlerdir. Eksik veri çözümlenmesi bu çalışmalardan sonra daha fazla ilgi çeker ve üzerinde çalışılır hale gelmiştir. Süreç içerisinde eksik veri yapısı ve düzenekleri varsayımlarda kullanılırken karıştırıldıkları ya da aynı olayı anlatıyormuş gibi yanlış anlaşılmalara sebep olmuştur.

Enders (2010), eksik veri yapısı ve eksik veri düzenekleri ile ilgili farkı ayırt etmenin gerektiğini belirtmiştir. Eksik veri yapısı, gözlenmiş değerler ile eksik değerlerin veri kümesi üzerinde nasıl bir düzene sahip olduğu ile alakalıdır. Eksik veri düzenekleri ise eksik veri olasılığı ve elde edilmiş değerler arasındaki mümkün ilişkileri tanımlar. Eksik veri yapısı kaybın neden olduğunu açıklamaz, sadece verideki boş gözlemlerin nerede olduğunu gösterir. Eksik veri düzenekleri ise verinin neden eksik olduğuyla ilgili

kesin bir açıklama veremese de veri kümesi ve eksik değer arasındaki matematiksel eşitliği gösterir.

Araştırmalarda eksik veri sorununun ancak 1970 sonrasında tartışılmaya başlanması ve eksik veri yöntemlerinin ancak 1990 sonrasında istatistiksel yazılımlara yansması, geliştirilmiş birçok eksik veri yönteminin henüz beklentileri karşılayacak düzeyde olmamasını bir ölçüde açıklamaktadır (Demir, 2013).

2.4. Eksik Veri Mekanizmaları (Türleri)

Eksik veriler veya eksik veri mekanizmaları genellikle üç kategoriye ayrılabilir; Tamamen Rastgele Eksik Olan Veriler (Missing Completely at Random-MCAR), Rastgele Eksik Olan Veriler (Missing at Random -MAR) ve Rastgele Eksik Olmayan Veriler (Missing Not at Random- MNAR). Bu eksik veri türlerinin ilk ikisi, dikkate alınamayan bir eksik veri mekanizmasından kaynaklanırken, sonuncusu eksik olmayan bir veri mekanizmasından kaynaklanan eksik veri olarak bilinir (Schafer, 1997; Allison, 2009). İhmal edilebilir bir mekanizmadan oluşturulan verilerin eksik olması, geçerli parametre tahminlerinin, eksik veri mekanizmasını modellemek zorunda kalmadan tahmin edilmesini sağlar. Diğer taraftan, dikkate alınamayan bir eksik veri mekanizması tarafından üretilen eksik veriler, geçerli parametre tahminlerinin elde edilebilmesi için genellikle mekanizmanın kendisinin bir modelini gerektirir. Bu nedenle, verilerde eksikliğe neden olan mekanizmanın anlaşılması veya varsayımlarda bulunulması eksik verilerin işlenmesinde doğru yöntemi seçmek için önemlidir.

Eksik veri mekanizmasının ihmal edilebilir olması, eksik verilerin analiz dışı bırakılabilmesi ile araştırmacının işini oldukça kolaylaştıracaktır. Fakat aksi durumda eksik veri mekanizmasının modellenmesi gerekir (Demir, 2013).

Eksik veri yapısı, mekanizma çeşitlerinden hangisine uyuyorsa yöntem, mekanizmanın izin verdiği koşullarda çözüme kavuşur. Aksi halde, eksik veri çalışması, mekanizma tarafından uygun olmayan yöntemle çözümlenebilir ve yanlış tahminler elde edilebilir. Herhangi bir eksik verinin nasıl oluştuğunu incelemeyen uygun olmayan eksik veri yöntemleri kullanımı çok iyi performans sağlamaz (Enders, 2010). Diğer bir deyişle, eksik veri mekanizmalarını göz önünde bulundurmadan yapılacak bir çalışmada daha fazla bilgi elde edebilmek adına kullanılan eksik veri çözümlemesi deyim yerindeyse

“kaş yapayım derken göz çıkartmak” gibi olacaktır. Bu yüzden eksik veri mekanizmalarının varsayımlarına göre çözümler yapmak gerekmektedir.

2.4.1. Tamamen rastgele eksik veri (missing completely at random- MCAR)

X ve Y bir veride, yaş ve cinsiyette olduğu gibi birbirinden bağımsız ve farklı değişkenler olsun. X ve Y'nin gözlenmesi ihtimali birbirinden etkilenmeyen değişkenler olduğu durumda X veya Y değişkeninin eksik veriye sahip olması tamamen rastgele eksik (MCAR) veri durumunu meydana getirmektedir (Allison, 2009; Enders, 2010).

Herhangi bir değişkendeki eksik veri olasılığı kendisine ve ölçülen diğer değişkenlere bağlı değildir. Başka bir ifadeyle gözlenen veri noktaları verilen cevapların basit rastgele örneklemevidir (Enders, 2010).

Allison (2009) ve King ve ark. (2001) eksik bir veri mekanizması hakkında yapılabilecek en güçlü ve çoğu zaman en gerçekçi olmayan varsayımın, mekanizmanın eksik verilerin tamamen rastgele (MCAR) eksik olmasına, yani tüm verilerin eksik olma olasılığına sahip olması olduğunu vurgulamışlardır. Bunun, bir verinin eksik olma ihtimalinin hem değişkenlerden hem de veri setinde yer alan diğer tüm değişkenlerden bağımsız olduğunu gösterdiğini belirtmişlerdir.

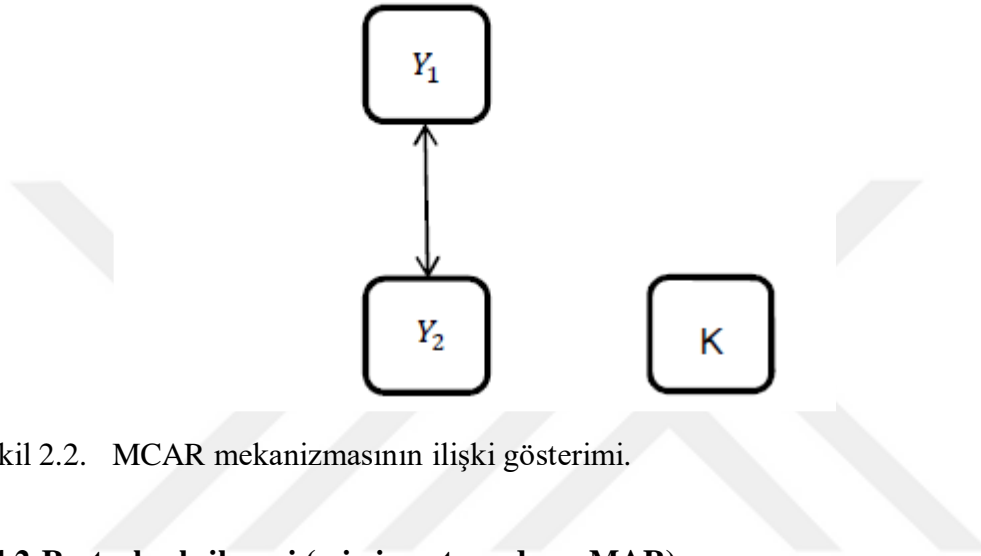
Tamamen rastgele eksiklik, bir değişkene ilişkin verinin eksik olma olasılığının araştırma kapsamında incelenen diğer değişkenlerden ve bu değişkenin kendisinden bağımsız olma durumudur. Bu durumda, X ve Y değişkenleri üzerinde yürütülen örnek bir araştırma için $P(X_{\text{eksik}} | X, Y) = P(X_{\text{eksik}})$ şeklinde ifade edilebilir. Bu ifadeye göre, X değişkenine ilişkin bir verinin eksik olma olasılığı X ve Y değişkenlerinin değerlerinden bağımsızdır (Akbaş ve Tavşancıl, 2015).

MCAR mekanizması araştırmacıların eksik veri çözümlerinde kolaylarına gelen bir mekanizmadır. Ama aynı zamanda varsayımları göz önünde bulundurulmadan yapılan çalışmalarda araştırmacıları yanlış çözümlere götürebilir. MCAR mekanizmasında herhangi bir eksik veri, ne diğer değişkenlerle ne de olduğu değişken ile ilişkilidir. Bu yüzden çok güçlü bir varsayımdır.

Donders ve ark. (2006), verinin MCAR olduğu bilindiğinde, eksik veri içermeyen değişkenler kümesinin, hedef kitlenin rastgele örnekleme olduğunu belirtmiştir. Bu

nedenle, eksik veri incelemeleri için MCAR varsayımında birçok yöntem yansız sonuçlar vermektedir (Satici, 2009).

İki başlı oklar istatistik birlikteliği anlatmak üzere, şekil 2.2. MCAR durumunun iki değişkenli veri kümesindeki görsel ifade biçimidir. Burada, K eksik veri gösterge matrisi, Y_2 değişkenindeki eksikler yardımıyla oluşturulsa K'nın istatistik ilişkisi bulunan herhangi bir değişken olmadığı görülebilir.



Şekil 2.2. MCAR mekanizmasının ilişki gösterimi.

2.4.2 Rastgele eksik veri (missing at random -MAR)

X ve Y gibi değişkenler arasından X'in bağımsız Y'nin ise bağımlı değişken olduğu durumlarda Y'nin cevap olasılığının X'e bağlı olmaması durumundaki eksik veri, rastgele eksik veri olarak adlandırılır. Başka bir ifadeyle Y değişkenindeki eksik verinin olasılığı analiz edilen modeldeki diğer değişken veya değişkenlerin ölçümlerine bağlı olduğu halde Y değişkeninin kendisine bağlı değilse verideki eksik veri mekanizması rastgele eksiktir (Enders, 2010).

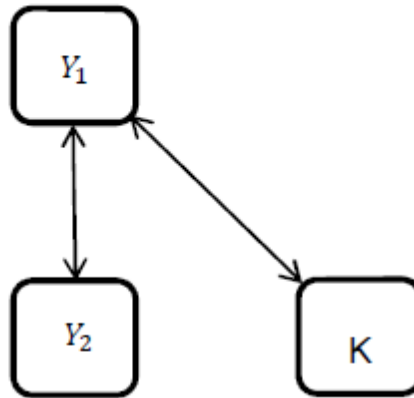
Daha az güçlü ve çoğu koşulda daha makul olan, eksik veri mekanizmasına ilişkin varsayım, mekanizmanın verilerin rastgele (MAR) olarak eksik olmasına neden olduğu varsayımdır. MAR varsayımlarına göre, verinin eksik olduğu değişkende belirli bir verinin eksik olması veya gözlenmesi olasılığı, verinin kendisinin değerinden bağımsızdır, ancak veri kümesinde yer alan diğer değişkenlere bağlı olabilir (Schafer, 1997; Allison, 2009).

Eksik veriyi oluşturan mekanizmayı veri setinde yer alan diğer değişkenlerle modellemek mümkün olduğundan, MAR olan eksik verilerin eksiklik mekanizmasının geçerli parametre tahminleri için ayrı ayrı modellenmesini gerektirmez.

Herhangi bir değişkende eksik veri olması olasılığı, modeldeki diğer değişkenlerle ilişkili ve kendi değişkenindeki gözlenen değerlerle ilişkili değilse eksik veri rastgele eksiktir. Enders (2010), MAR durumunun aslında eksik veri olasılığı ile bir veya daha fazla değişkenin arasındaki sistematik ilişkiyi açıkladığını belirtmiştir. Genellikle veri MAR olduğu durumda, veri yükleme gibi basit yöntemler yanlış sonuçlar vermektedir. Bununla birlikte, model tabanlı veri yükleme ve çoklu veri yükleme gibi daha ayrıntılı yöntemler yansız sonuçlar vermektedir (Donders ve ark., 2006; Satıcı, 2009).

Rastgele eksik, herhangi bir verinin eksik olma olasılığının kaybın bulunduğu değişkenin değerinden bağımsız, fakat araştırma kapsamında ölçülen diğer bir değişkene (ya da değişkenlere) bağımlı olması durumudur. $P(X_{\text{eksik}} | X, Y) = P(X_{\text{eksik}} | Y)$ şeklinde ifade edilebilecek olan bu durum, X değişkenindeki bir verinin eksik olma olasılığının Y değişkeninin değerlerine bağlı olduğu anlamına gelmektedir (Akbaş ve Tavşancıl, 2015).

İki başlıklı oklar istatistik birlikteliği anlatmak üzere, Şekil 2.3. iki değişkenli durum için MAR durumunun görsel ifade biçimidir. K eksik veri gösterge matrisi, Y_2 değişkenindeki eksiklerden oluşursa. K'nın istatistiksel ilişkisi bulunan değişkenin sadece Y_1 olduğu görülebilir.



Şekil 2.3. MAR mekanizmasının ilişki gösterimi.

MCAR'ın, MAR'dan daha kısıtlayıcı bir eksik veri düzeneği olduğu görülmelidir. MCAR durumunda eksik veri olasılığı herhangi bir değişkene bağlı değilken, MAR durumunda eksik veri olasılığının diğer değişkenlerle ilişkili olduğu görülmektedir. MAR, ihmal edilebilir eksik olarak düşünülerek çeşitli çözümlerle sıkça kullanılmıştır. Rubin (1976) ise örnekleme dayalı olarak yapılan çözümler yöntemlerinin MCAR olması durumunda geçerli olabileceğini göstermiştir.

Eksik verinin MCAR varsayımını sağlaması geleneksel yöntemler için uygundur. Ancak MCAR'nin kısıtlayıcı, zor bir varsayım olması ve model tabanlı veri yükleme yöntemlerinin önerilmesi, geleneksel yöntemlerin terkedilmesine neden olmuştur. Bu süreçte MCAR ve MAR varsayımları altında yansız kestirimler elde edebilen beklenti maksimizasyon kestirimleri ve çoklu veri yükleme yöntemlerinin rolü büyüktür (Wilkinson, 1999).

2.4.3. Rastgele olmayan eksik veri (missing not at random - MNAR)

Herhangi bir değişkendeki eksik veri ihtimali değişkenin kendisi ile ilgili olduğu halde diğerleriyle alakalı değilse bu şekildeki eksik veri mekanizmasına Rastgele Olmayan Eksik Veri (MNAR) denir. Veri setindeki eksik veri, yukarıdaki varsayımların hiçbiri için geçerli kabul edilemezse, eksik veri mekanizmasının verilerin rastgele (MNAR) eksik olmaması olduğu kabul edilir. Temel olarak, bu, Y için belirli bir değerin eksik olup olmadığı veya gözlenen değerin, Y'nin değerine bağlı olduğu ve bu değeri, gözlemlenen X değişken kümesinin değerlerinden tahmin etmenin mümkün olmadığı anlamına gelir (King ve ark., 2001; Allison, 2009).

Seçim yanlılığını gidermek için kullanılan yöntemler MNAR olan eksik verilerle başa çıkmak için kullanılan birkaç farklı yöntem arasındadırlar (King ve ark. 2001). Bununla birlikte, bu yöntemler eksik veri mekanizmasının analizini gerektirir, çünkü her MNAR durumu benzersizdir ve bu nedenle MNAR verilerine genelleştirilebilir yaklaşımlar oluşturmak zordur. Ayrıca, MAR'da görüldüğü gibi, veri setine daha fazla değişken ekleyerek MNAR durumunu iyileştirmek genellikle mümkündür. Bu mutlak biçimde MNAR eksik veriyi MAR eksik veriye dönüştüremeyebilir, ancak MNAR probleminin ciddiyetini azaltabilir.

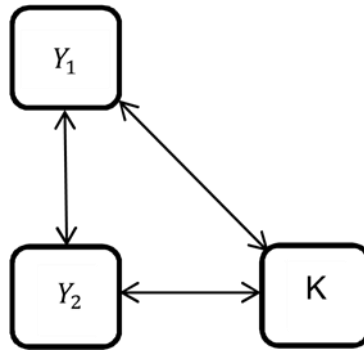
Herhangi bir deęişkendeki eksik verinin olasılığı modeldeki dięer deęişkenlerle ve aynı zamanda kendi deęişkeniyle de ilişkiliyse eksik veri MNAR'dır. MCAR ve MAR olmayan verinin MNAR olduęu da söylenebilir. MNAR mekanizmaları altında eksik veri çözümlenmesi mevcut yöntemleri kullanmaya uygun deęildir.

Rastgele olmayan eksik, herhangi bir verinin eksik olma olasılığının kaybın bulunduğu deęişkenin deęerine baęlı olması durumudur. $P(X_{\text{eksik}} | X, Y) = P(X_{\text{eksik}} | X)$ şeklinde ifade edilebilecek olan bu durum, X deęişkenindeki bir verinin eksik olma olasılığının kendi deęerine baęlı olduęu anlamına gelmektedir (Rubin, 1976; Little ve Rubin, 1987).

Eksik verinin uyduęu düzeneęe göre veri silme, veri yükleme veya model tabanlı yöntemleri kullanma durumları ortaya çıkar. Eksik veri çözümlenmesi için kullanılan yöntemler ileride daha detaylı anlatılacak ve eksik veri mekanizmaları ile olan ilişkileri açıklanacaktır. Burada çözümlenme yöntemlerine deęinme amacı, MCAR, MAR ve MNAR ile ilgili örneklerin açıklanabilmesidir.

MCAR veri mekanizmasında liste boyunca silme, ikili silme ya da veri yükleme yöntemleri kullanılabilir. MAR veri düzeneğinin varsayımlarından dolayı veri yükleme yöntemleri yanlı sonuç verirken model tabanlı yöntemlerdeki sonuçlar daha etkindir.

Şekil 2.4. , MNAR durumunun iki deęişken bulunduğu durumda görsel ifade biçimidir. Şekil incelendiğinde, K eksik veri gösterge matrisi, Y_2 deęişkenindeki kayıplardan oluşturulsun. K'nın istatistiksel ilişkisi bulunan deęişkenlerin Y_1 ve Y_2 olduęu görülebilir.



Şekil 2.4. MNAR mekanizmasının ilişki gösterimi.

Konuyu örneklendirme adına Enders (2010)'in çalışmasındaki örnek üzerine çalışılmıştır. Bireylerin zekâ düzeyleri ve aynı bireylerin belirli bir dönem için gösterdikleri iş performans puanları ile ilgili bilgiler Çizelge 2.1'deki gibi elde edilmiştir. İş performans puanları, MCAR, MAR ve MNAR yapılarına uygun olacak şekilde oluşturulmuş ve akabinde tartışmalar yapılmıştır. Her bir eksik veri düzeneği ayrı ayrı konularmış gibi incelenmiştir.

Çizelge 2.1. Zekâ puanı ile MCAR, MAR ve MNAR mekanizmaları oluşturulmuş iş performans puanları

Sıra	Zekâ Puanı	İş Performans Puanları			
		Tam Veri	MCAR	MAR	MNAR
1	78	9	—	—	9
2	84	13	13	—	13
3	84	10	—	—	10
4	85	8	8	—	—
5	87	7	7	—	—
6	91	7	7	7	—
7	92	9	9	9	9
8	94	9	9	9	9
9	94	11	11	11	11
10	96	7	—	7	—
11	99	7	7	7	—
12	105	10	10	10	10
13	105	11	11	11	11
14	106	15	15	15	15
15	108	10	10	10	10
16	112	10	—	10	10

Çizelge 2.1. Zekâ puanı ile MCAR, MAR VE MNAR mekanizmaları oluşturulmuş iş performans puanları (devamı)

Sıra	Zekâ Puanı	Tam Veri	İş Performans Puanları		
			MCAR	MAR	MNAR
7	113	12	12	12	12
18	115	14	14	14	14
19	118	16	16	16	16
20	134	12	—	12	12

İlk olarak, 5 tane iş performans puan değeri MCAR olacak şekilde rastgele silinmiştir. Silinmiş olan bu 5 adet veri ne iş performansı değişkeniyle ne de zekâ düzeyi ile ilişkilidir. Bu şirketin iş performansı ile ilgili analizler, geriye kalan 15 veri üzerinden elde edilebilir.

Aynı örnek şirkette işe alınacak olan 20 kişi için işe alma kararının deneme süresindeki iş performansları dikkate alınmadan sadece zekâ puanlarına göre verildiği öngörülün. Oluşturulan eksik veri zekâ puanı en düşük 5 kişiden oluşturulsun. Oluşan eksik veri MAR olacak şekilde elde edilmiştir. Eksik veri yapısı MAR olduğundan eksik iş performans puanları kendi değişkenlerindeki değerlere değil; zekâ düzeyi değişkenine bağlı olacaktır.

Son olarak, işyerindeki 20 kişiden iş performansı en düşük olan 5 kişinin işine son verileceği düşünülün. Oluşan eksik veri yapısı MNAR eksik veri yapısına uygun olacaktır.

MCAR eksik veri mekanizması, mekanizmalar içinde test edilebilen tek mekanizmadır. Bu mekanizma Little'in MCAR Testi tarafından test edilebilmektedir.

Eksik veri mekanizmaları MCAR ile MAR arasında bazı farklılıklar vardır. Bu farklılıklar yapılacak analizlerde önemli etkilere sahip olmaktadır. Toka (2012), veri mekanizmasının MCAR veya MAR olması arasındaki farklılıkları genel haliyle Çizelge 2.2.'deki gibi göstermiştir.

Çizelge 2.2. MCAR ve MAR mekanizmaları arasındaki farklılıklar (Toka, 2012)

MCAR	MAR
MCAR mekanizmalar içerisinde en kısıtlayıcı olanıdır.	MCAR'dan daha az kısıtlayıcıdır.
MCAR mekanizmaları geleneksel veri yükleme yöntemlerinde yansız sonuçlar verebilmektedir.	MAR, ML kestirim ve çoklu atama gibi yöntemlerde yansız sonuçlar verebilmektedir.
MCAR, Little tarafından geliştirilen bir testle test edilebilir bir düzendir (Little'ın MCAR Testi).	MAR için test edilebilir bir yöntem geliştirilememiştir.
MCAR, kısıtlayıcı özelliğinden dolayı birçok eksik veri düzeneğinde sağlanamaz.	MAR eksik veri düzeneğini elde etmek daha kolaydır.

Eksik veri mekanizmaları olan ve yukarıda anlatılan MCAR, MAR ve MNAR terimlerine kısa bir göz atalım. Rubin (1976) eksik veri problemlerini üç kategoride sınıflandırmıştır. Teorisinde her veri noktasının eksik olma olasılığı vardır. Bu olasılıkları yöneten sürece eksik veri mekanizması ya da cevap mekanizması denir. Sürecin modeli eksik veri modeli veya yanıt modeli olarak adlandırılır.

Eksik olma olasılığı tüm durumlar için aynıysa, verilerin rastgele (MCAR) tamamen eksik olduğu söylenir. Bu, eksik verilerin nedenlerinin verilerle ilgisi olmadığını gösterir. Sonuç olarak, bariz bilgi kaybının yanı sıra, veriler eksik olduğu için ortaya çıkan karmaşıklıkların çoğunu görmezden gelebiliriz. MCAR örneği, bataryaları tükenmiş bir terazi gibidir. Verilerin bir kısmı sadece şanssızlık nedeniyle kaybolacaktır. (Enders, 2010).

Eksik olma olasılığı sadece gözlenen veriler tarafından tanımlanan gruplar içinde aynıysa, o zaman veriler rastgele (MAR) olarak eksiktir. MAR, MCAR'dan çok daha geniş bir sınıftır. Örneğin, yumuşak bir yüzeye yerleştirildiğinde, tartım terazisi sert bir yüzeye yerleştirilenden daha fazla eksik değer üretebilir. Bu nedenle bu veriler MCAR değildir.

Bununla birlikte, yüzey tipini biliyorsak ve MCAR'ı yüzey tipinde varsayabilirsek, o zaman veriler MAR'dır. MAR, MCAR'dan daha genel ve daha gerçekçidir. Modern eksik veri yöntemleri genellikle MAR varsayımından başlar (Enders, 2010).

MNAR yapısındaki eksik verilerde sistematik bir eksiklik söz konusu olduğundan bu eksiklikler ihmal edilerek yapılacak tahminler sapmalı olacaktır (Silva ve Zarate, 2014). Bu yüzden MNAR yapısındaki eksik veriler “ihmal edilemez” eksik olarak nitelendirilmektedir. Bu yapılar için genel bir örnek, Little ve Rubin tarafından gelir ve yaş değişkenleri üzerinden şu şekilde verilmiştir; gelir değişkeninde eksikliklerin olması kişinin yaşı veya geliri ile ilişkili değilse MCAR, kişinin yaşı ile ilişkili, ancak geliri ile ilişkili değilse MAR, kişinin geliri ile ilişkili ise MNAR mekanizması geçerlidir (Little, ve Rubin, 1987).

Hem MCAR hem de MAR 'un geçerli olmadığı durumlarda, MNAR'dan söz edebiliriz. Literatürde aynı kavram için NMAR (Not Missing at Random) terimi de kullanılır. MNAR, eksik olma olasılığının bizim için bilinmeyen nedenlerden dolayı değiştiği anlamına gelir. Örneğin, terazi ölçek mekanizması zaman içinde yıpranabilir, zaman ilerledikçe daha fazla eksik veri üretebilir, ancak bunu not etmekte başarısız olabiliriz. Daha ağır nesnelere zaman içinde ölçülürse, o zaman çarpılacak ölçümlerin bir dağılımını elde ederiz. MNAR, ölçeğin daha ağır nesnelere için (yukarıdaki gibi) daha fazla eksik değer üretme olasılığını içerir; bu, tanınması ve kullanılması zor bir durumdur. Kamuoyu araştırmalarında MNAR örneği, zayıf görüşleri olanların daha az yanıt vermesi durumunda ortaya çıkar. MNAR en karmaşık durumdur. MNAR ile başa çıkma stratejileri, eksikliğin nedenleri hakkında daha fazla veri bulmak veya sonuçların çeşitli senaryolar altında ne kadar hassas olduğunu görmek için eğer-eğer analizleri yapmaktır (Allison, 2000; Enders, 2010).

Rubin'in ayrımı, bazı yöntemlerin neden işe yarayıp yaramayacağını anlamak için önemlidir. Teorisi, eksik bir veri yönteminin geçerli istatistiksel çıkarımlar sağlayabileceği koşulları ortaya koymaktadır. Çoğu basit düzeltmeler yalnızca kısıtlayıcı ve çoğu zaman gerçekçi olmayan MCAR varsayımı altında çalışır. Eğer MCAR uygun değilse, bu tür yöntemler yanlış tahminler sağlayabilir (Buuren, 2012).

3. EKSİK VERİLERİ YÖNETME YÖNTEMLERİ

Giriş bölümünde bahsedildiği gibi, eksik veriler, çoğu araştırma alanında bir biçimde ortaya çıkan bir sorundur. Ancak, eksik verilerin ele alınmasına ihtiyaç duyulmasına rağmen, eksik veri problemi her zaman tartışılmayabilir ve bu problemi çözmek için hangi yöntemlerin kullanılacağı bazen belirsizdir (Buuren, 2012). Eksik verinin nasıl oluştuğunu tartışmamak cazip gelebilir, çünkü tartışmama verilerin kalitesi ve çalışmanın kalitesiyle ilgili sorulardan kaçınmayı mümkün kılabilir. Bununla birlikte, bu tartışmadan kaçınmak, altta yatan problemi tespit etmeden eksik verileri ele almak için uygun olmayan yöntemlerin kullanılmasına yol açabilir.

Eksik verilerin ele alınmasında kullanılan yöntemlerin değerlendirilebileceği üç genel teorik kriter vardır. İlk olarak, eksik değerler gözlemlenirse eksik veri yöntemi analizde parametre yanlılığını en aza indirmeli, yani işlenmiş veriler üzerinde istatistiksel analiz yapıldığında, eksik veri yöntemi, ilgilenilen parametrelerin elde edilecek değerlere mümkün olduğunca yakın olmasını sağlamalıdır (Buuren, 2012). İkincisi, eksik veri yöntemi, mevcut verilerin maksimum miktarının kullanılmasını ve verilerin mümkün olduğunca azının atılmasını sağlamalıdır. Çünkü veri toplama genellikle zaman alıcı ve pahalı bir iştir. Son olarak, eksik veri yönteminin, verilerin değişkenliği hakkında iyi tahminler vermesi, yani ne büyük ne de çok küçük olmayan standart hatalar, p değerleri ve güven aralıkları üretmesi gerekir. Allison (2009) bu son ölçütün, analizden elde edilen çıkarım için, standart hataların çok büyük ya da çok küçük tahmin edilmesi ve dolayısıyla p değerlerinin ve güven aralıklarının uzunluğunun, standart hataların ve hesapların uzunluğunun hafife alınmasında analizde daha düşük verime yol açması bakımından önemli olduğunu vurgulamıştır.

Eksik verilerin varlığında, kullanılabilir olan birçok istatistiksel yöntem geliştirilmiştir. Kullanılacak iyi bir yöntemin, eksik verilerden kaynaklanabilecek yanlılığı minimize etmesi, elverişli bilgilerin kullanımını maksimize etmesi ve belirsizliklere yönelik standart hata, güven aralığı ve olasılık değerlerine yönelik titiz kestirimler sağlaması beklenir (Allison, 2009).

Araştırmalarda eksik veri sorunu çok yaygın olduğundan, yalnızca eksik verileri ele almak için kullanılan yöntemlerin teorik performans açısından değil, aynı zamanda

kullanım kolaylığı ve hesaplama etkinlikleri açısından değerlendirilmesi önemli bir noktadır. Bu iki boyut önemlidir çünkü eksik veri yöntemlerinin kullanımı kolay ve hızlı değilse, sınırlı istatistik bilgisine sahip araştırmacılar, eksik verileri ele almak için daha az karmaşık diğer yöntemleri kullanmayı tercih edebilir.

Eksik verilere yönelik ilk akla gelen ve alışıldık yöntem, eksik verilerin analiz dışı bırakılması ya da eksik veriler yerine basit veri atama yöntemidir.

Silmeye dayalı yöntemler içerisinde en bilindik olanlar 'dizin silme-DS (listwise deletion)', ve 'kısmî silme-KS (pairwise deletion)' yöntemleridir. Basit atamaya dayalı yöntemler içerisinde en bilindik olanı ise 'ortalama atama-OA (mean substitution)' yöntemidir (Demir, 2013).

Araştırmalarda karşılaşılan eksik veri sorununun çözümünde kullanılmak üzere geçmişten günümüze kadar değişik yöntemler geliştirilmiştir. Eksik veri ile analizi sürdürme, eksik gözlemler yerine veri atama veya çeşitli istatistiksel yöntemlerle eksik verileri tamamlama gibi yöntemler bu yöntemlerden bazılarıdır (Little, 1988; Downey ve King, 1998; Duncan, Duncan ve Li, 1998; Bal, 2003; Carpita ve Manisera, 2011). Bu yöntemler içerisinde araştırmacılar tarafından en çok kullanılan yöntemler liste boyunca silme ve ikili silme gibi eksik verileri analizin dışında bırakma yöntemleridir. Ancak yapılan çalışmalar bu yöntemlerin örnekleme kaybı, güvenilirlikte azalmaya, tahminlerde yanlılığa neden olduğunu (Allison, 2009; Satıcı ve Kadılar, 2009; Van der Ark ve Vermunt, 2010; Cumming, 2013) ve yanlılıktan kaynaklı olarak da örneklemin evreni temsil etme derecesinin düştüğünü göstermektedir (Little, 1988; Demir ve Parlak, 2012). Belirtilen bu sebeplerden dolayı, son yıllarda, bu yöntemler yerine, beklenti maksimizasyonu ve çoklu atama gibi modern yöntemler önerilmektedir (Kürşad ve Nartgün, 2015).

Veri kümelerinde var olan eksik verilerle baş edebilmek için eksik verili kaydı silmek, eksik veriyi göz ardı etmek, sık geçen ifade ile doldurmak veya ilgili kayıt ya da niteliği satır, sütun ortalaması ile doldurmak gibi basit klasik yöntemler karmaşık hesaplama yöntemlerinin yerine kullanılmaktadır (Cheng ve Victoria-Feser, 2011; Meng ve Shi, 2012). Ancak bu gibi basit klasik yöntemler eksik verileri yok sayarak verimi düşürmekte bununla beraber var olan veriyi yanlışlaştırarak sistematik anlamda kalitesizleştirmektedir (Little ve Rubin, 1987; Pelckmans ve ark. 2005; Nelwamondo ve Marwala, 2008).

Arařtırmacıların eksik veri problemi çözümünde günümüze kadar kullandığı yöntemler incelendiğinde yöntemlerin en genel anlamda geleneksel yöntemler ve modern atama yöntemleri olarak iki alt başlıkta ele alınabileceği görülür.

3.1. Geleneksel Yöntemler

Eksik verilerle başa çıkmak için kullanılan en yaygın yöntemler, eksik değerleri veya eksik verilerle yapılan tüm gözlemleri kaldıran yöntemler olan silme yöntemleri ile basit atama yöntemleridir. Aşağıda gösterildiği gibi, eksik verileri ele alan bu "geleneksel" yöntemler, özellikle veriler MCAR olmadığında, yukarıdaki üç teorik kriter üzerinde iyi performans göstermez.

3.1.1. Liste boyunca silme (listwise deletion -dizin silme)

Çoğu istatistiksel yazılımdaki ve muhtemelen çoğu arařtırmadaki eksik verilerle başa çıkmanın standart yaklaşımı, sadece eksik verileri olan tüm gözlemleri silmektir (Buuren, 2012). Eksik verilerle başa çıkma konusundaki bu yaklaşım, tüm değişkenlerin gözlemlendiği gözlemleri veya vakaları kullandığı için listwise deletion veya complete case analysis olarak adlandırılır (King ve ark., 2001; Allison, 2009). Liste boyunca silme, herhangi bir hesaplama veya manipölasyon gerektirmedikten eksik verileri ele almak için çok basit bir yöntemdir ve çoğu istatistiksel analiz yazılımının, eksik verilerle yapılan çalışmalarda, eksik gözlemleri analiz dışı tutma varsayımlarına dayanan seçenekleri vardır. Oldukça kolay olan liste boyunca silme, ilk bakışta pratik ve basit bir çözüm gibi durmaktadır. Bununla birlikte liste boyunca silme, temelde bazı verilerin ihmal edilmesine yönelik bir yaklaşımdır ve veri kaybı söz konusudur. MCAR varsayımı altında, liste boyunca silme, regresyon ayarında olduğu gibi ortalamalar, varyanslar için de tarafsız parametre tahminleri verir. Bununla birlikte, gözlemler çıkarıldığında numune boyutu küçülecek ve liste boyunca silme kullanılarak hesaplanan regresyonlardan üretilen standart hatalar, eksik veriler analize dâhil edilerek üretilen standart hatalardan daha büyük olacaktır (Buuren, 2012).

Veriler MAR ise liste boyunca silme artık eksikliğin bağımlı olduğu değişkenler arasındaki ilişkinin ve eksikliğin olduğu değişkenin parametrelerde yanlılığa yol

açacaktır. Ayrıca aralarındaki korelasyon için yansız parametre tahminleri vermeyecektir. Örneğin, ilgilenilen parametre bir değişkenin ortalamasıysa ve değişken birinci grup ortalamasının ikinci grubun ortalamasının iki katı olduğu iki grupta ölçülürse ve daha sonra, ilk grubun değişkenle ilgili bir soruyu ikinci gruba göre daha fazla yanıtlaması muhtemelse, ortalamayı hesaplamak için liste boyunca silme kullanılmışsa, bu durum için yukarı doğru önyargı yaratacaktır (Allison, 2009).

Dezavantajlarına rağmen liste boyunca silmenin MCAR ve MAR varsayımlarına karşı oldukça kararlı ve sağlam olduğu ifade edilmektedir. Bu durum özellikle regresyon analizlerinde önemli bir üstünlük sağlamaktadır. Liste boyunca silme yönteminin bir diğer olumlu özelliği de, gerçek standart hata kestirimindeki titizliğidir. Bu ve benzeri olumlu yanları nedeniyle liste boyunca silme, geleneksel eksik veri yöntemleri arasında ‘iyi’ bir yöntem olarak değerlendirilmektedir (Little ve Rubin, 1987; Allison, 2009).

Enders (2010) liste boyunca silme yönteminde, herhangi bir değişkene ait bir verinin eksik olması durumunda, eksik veri içeren tüm gözlemlerin veri setinden çıkarıldığını vurgulamıştır. Ayrıca liste boyunca silme yönteminin uygulanmasıyla tam bir veri setine ulaşıldığı için her türlü istatistiksel analizin gerçekleştirilebileceğini göstermiştir.

Liste boyunca silmeye örnek bir veri tabanı olarak Çizelge 3.1.’i ele alalım.

Çizelge 3.1. Liste silme için örnek veri tabanı

Satır No	İsim	Puan	Ders
1	A	70	Matematik
2	B	55	İngilizce
3	C	-	Bilgisayar
4	D	80	-
5	E	30	Fizik
6	-	-	Matematik

Çizelge 3.1.’de 3,4 ve 6.satırlarda eksik bilgi bulunmaktadır. Bu durumda, örneğin öğrencilerin puan ortalamalarının alınması istendiğinde hata ile karşılaşılacaktır.

Liste boyunca silme yaklaşımında bu eksik veri içeren satırların tamamı tablodan temizlenir ve veri tabanındaki tablo, Çizelge 3.2.’ de gösterilen hale getirilir.

Çizelge 3.2. Liste silme sonrası örnek veri tabanı

Satır No	İsim	Puan	Ders
1	A	70	Matematik
2	B	55	İngilizce
5	E	30	Fizik

Çizelge 3.2.'deki veri tablosunda herhangi bir eksik veri bulunmamaktadır ve böyle bir veri seti ile çalışıldığında eksik veriden kaynaklanan herhangi bir sorunla karşılaşılması beklenir. Liste boyunca silme yöntemi, veri tabanındaki faydalı olabilecek diğer verilerin kaybolmasına sebep olduğu için dezavantajları olan bir yöntemdir. Örneğin herhangi bir sütunda çok sayıda eksik veri bulunması durumunda (ki bu eksik veri oranı Çizelge 3.1.'de %50 'dir), veri tabanının yorumlanmasını önemli bir ölçüde etkilemektedir.

Rubin (2002), liste silme işleminde gerçekleşen bilgi kaybının kesinliği etkileyeceğini ve eksik veri düzenek varsayımının MCAR olmaması durumunda yanlış kestirimler elde edilebileceğini belirtmiştir. Ayrıca Baygül (2007), liste bazında veri silme yöntemi kullanılırken örneklemdaki gözlem sayısı azaltılacağından standart hataların artacağını belirtmiştir.

Özetle, Liste Boyunca Silme, eksik verileri ele almak için üç kriterden hiç birini iyi karşılamamaktadır, çünkü yansız parametre tahminleri üretmemektedir (veri MCAR olmadığı sürece), çok fazla toplanmış bilgiyi boşa harcar ve belirsizliğin kesin tahminlerini vermez. Teorik eksikliklerine rağmen liste boyunca silme çok kolay bir yöntemdir. Ek hesaplamalar gerektirmediğinden, eksik verilerle başa çıkmak için kullanılan standart yöntemlerden biridir (Buuren, 2012).

Eksiksiz durum analizi (liste boyunca silme), eksik verileri SPSS, SAS ve Stata dâhil olmak üzere birçok istatistiksel pakette ele almanın varsayılan yöntemlerinden biridir. İşlem, analiz değişkenlerinde bir veya daha fazla eksik değeri olan tüm durumları ortadan kaldırır. Tam vaka analizinin önemli bir avantajı kolay olmasıdır. Veriler MCAR ise, liste boyunca silme, ortalamaların, varyansların ve regresyon ağırlıklarının yansız tahminlerini üretir (Buuren, 2012). MCAR altında, liste boyunca silme, verilerin

azaltılmış alt kümesi için doğru olan ancak çoğu zaman mevcut tüm verilere göre daha büyük olan standart hatalar ve önem düzeyleri oluşturur.

Liste boyunca silmenin bir dezavantajı, potansiyel olarak israf edici olmasıdır. Gerçek hayattaki uygulamalarda, gerçek verinin yarısından fazlasının kaybedilmesi, özellikle değişken sayısı yüksek değilse olağandır. King ve ark (2001), siyaset bilimlerindeki eksik kayıtların yüzdesinin ortalama olarak %50'yi aştığını, bazı çalışmaların ise %90'ın üzerinde tamamlanmamış kayıtlara sahip olduğunu tespit etmişlerdir. Daha küçük bir alt örneğin ilgi etkilerini saptama yeteneğini ciddi şekilde bozabileceği açıktır.

Veriler MCAR değilse, liste boyunca silme, ortalama, regresyon katsayıları ve korelasyon tahminlerini ciddi şekilde önyargılı hale getirebilir. Little ve Rubin (2002), tahmini ortalamadaki yanlılığın, gözlemlenen ve eksik vakalar arasındaki fark ile ve eksik verilerin oranı ile arttığını göstermişlerdir.

Schafer ve Graham (2002), MAR ve MNAR'da liste boyunca silme eğilimini gösteren zarif bir benzetim çalışması yapmışlardır. Liste boyunca silme her zaman kötü sonuçlar göstermeyeceğini ifade etmişlerdir. Eksik verinin sonuçları, nerede olduklarına (sonuçlar veya tahminler) ve tüm veri modelinin parametre ve model biçimine bağlı olarak farklılık gösterir. Regresyon analizi bağlamında, liste boyunca silme, belirli ortamlarda onu çekici kılan bazı benzersiz özelliklere sahiptir (Buuren, 2012).

Eksik veri sorunu, örneğin veri setinin sadece küçük bir kısmı atılarak çözülebilirse, yöntem oldukça etkili olabilir.

3.1.2. Çift yönlü silme (pairwise deletion-ikili silme)

Eksik değerler içeren tüm vakaların kaldırılmasının bir alternatifi, istatistiksel analizde ilgilenilen çoğu parametrenin, araçların ve kovaryans matrisinin işlevleri olarak kullanılabilineceği ifade edilmektedir. Ortalamalar ve kovaryans matrisi daha sonra, her durum için mevcut veriler kullanılarak tahmin edilebilir. Böylece, bireysel değişkenler için ortalamalar ve varyanslar, o değişken için gözlemlenen veriler kullanılarak tahmin edilir. Ayrıca değişkenler arasındaki kovaryanslar, her iki değişken üzerindeki verilerle birlikte tüm gözlemler kullanılarak tahmin edilir. Bu eksik veri işleme yaklaşımına “Çift

Yönlü Silme veya Mevcut Vaka (Durum) Analizi” denir. Ayrıca liste boyunca silme'ye bir alternatiftir (Allison, 2009).

Mevcut durum analizi olarak da bilinen ikili silme, liste boyunca silme işleminin veri kaybı sorununu gidermeye çalışır. Yöntem, gözlemlenen tüm verilerdeki ortalamaları ve (ko)varyanslarını hesaplar. Bu nedenle, bağımsız değişkenin ortalaması, üzerinde gözlemlenen verilere sahip tüm durumlara dayanmaktadır (Buuren, 2012).

Çift yönlü silme ile elde edilen parametreler, MCAR varsayımı altında tutarlıdır; bu, asimptotik olarak yansız oldukları anlamına gelir. Veriler MAR olduğunda ve MCAR olmadığında, çift yönlü silme, liste boyunca silme gibi, yanlış tahminler verebilir. Bununla birlikte, her bir kovaryans modeli teorik olarak farklı sayıda gözlemle tahmin edilebileceği gibi, standart hataların çift yönlü silme ile regresyon ayarında tahmin edilmesi çok karmaşıktır. Standart hatalar örneklem büyüklüğünün bir fonksiyonu olduğu için bu hataları doğru şekilde tahmin etmek imkânsızdır. Çift yönlü silme ile ilgili ek bir komplikasyon da bu eksik veri yöntemi kullanılarak elde edilen kovaryans ve korelasyon matrislerinin, çoğu istatistiksel analiz yönteminde kullanım için bir gereklilik olan tüm değişkenlerin bulunmamasıdır (Allison, 2009; Graham, 2009).

İkili Silme, mevcut tüm bilgileri kullandığı için, teorik seviyede liste boyunca silmeye göre daha kolaydır. Diğer iki teorik değerlendirme kriterinde ise, İkili silme, liste boyunca silmeden daha iyi bir sonuç almaz. Ek olarak, hesaplama açısından yoğun bir hesaplama gerektirmese de, çift yönlü silme, sınırlı bir istatistik bilgisine sahip çoğu kullanıcı için, özellikle de standart hataların hesaplanması konusunda, çok kullanıcı dostu olarak değerlendirilmemektedir (Buuren, 2012).

Çift yönlü silme, ortalama ve standart sapma gibi dağılımı tanımlayıcı ölçülerin ve korelasyon ve kovaryans gibi ilişki ölçülerinin elde edilmesinde gözlenmiş veri çiftlerinin dikkate alınması esasına dayanır. Hesaplamalar her bir değişken ya da değişken çifti için eksiksiz veriye sahip bireyler üzerinden yapılır. Bu yöntem sayesinde, iki değişken arasındaki gözlem sayısını maksimize ederek korelasyon katsayılarını hesaplamak ve dolayısıyla örnekteki tüm ikili bilgiyi kullanmak mümkün olur (Alpar, 2017).

Çift silmeyi, liste boyunca silme için verdiğimiz örnek üzerinde çalışarak ele alalım.

Çizelge 3.3. Eşlerin silinmesi öncesi veri tabanı

Satır No	İsim	Puan	Ders
1	A	70	Matematik
2	B	55	İngilizce
3	C	-	Bilgisayar
4	D	80	-
5	E	30	Fizik
6	-	-	Matematik

Bu yöntemde tabloda eksik verilerin bulunduğu noktalara karşılık gelen bütün satırların temizlenmesi yerine sadece üzerinde işlem yapılan verilerdeki eksik veriler temizlenir. Çizelge 3.3.'de puan ortalamalarını hesaplamak istersek ve ikili silme işlemini uygularsak sadece puan sütununda eksik verinin olduğu satırlar silinir. Böylece 3 ve 6. satırlar silinir.

Çizelge 3.4. Eşlerin silinmesi sonrası veri tabanı

Satır No	İsim	Puan	Ders
1	A	70	Matematik
2	B	55	İngilizce
4	D	80	-
5	E	30	Fizik

Görüldüğü gibi sadece o değişken için problem çıkaran satırlar silinmiş ve diğer satırlar farklı sütunlarda eksik veri bulundurmasına rağmen ellenmemiştir.

Bu yöntem basittir ve mevcut tüm bilgileri kullanıyor gibi görünmektedir. Bu nedenle MCAR altında, ortalama, korelasyon ve kovaryansların tutarlı tahminlerini üretmektedir (Little ve Rubin, 2002).

Yöntemin bazı eksiklikleri de vardır. İlk olarak, veriler MCAR değilse tahminler önyargılı olabilir. Ayrıca, kovaryans ve/veya korelasyon matrisi çoğu çok değişkenli veri düzenekleri için gereklilik olan pozitif kesin(matrisin eksiksiz olması) olmayabilir.

Sorunlar genel olarak yüksek korelasyonlu deęişkenler için daha deęişkendir (Little 1988). Standart hataların hesaplanmasında hangi örneklem boyutunun kullanılması gerektięi açık deęildir. Ortalama örneklem büyüklüğünün alınması çok küçük standart hatalar verir (Küçük, 1992). Ayrıca, ikili silme, yaklaşık normal bir dağılım izleyen sayısal veriler gerektirir; oysa pratikte çoęu zaman karışık tipteki deęişkenlere sahibiz.

Mevcut tüm bilgileri kullanma fikri güzel olmakla beraber çiftli matrislerin doęru analizi, standart hataları hesaplamak için karmaşık optimizasyon yöntemleri ve özel formüller gerektirir (Praag ve Velzen, 1985; Marsh, 1998). İkili silme, veriler çok deęişkenli normal bir dağılıma yaklaşıyorsa, deęişkenler arasındaki korelasyonlar düşükse ve MCAR varsayımı geçerliyse kullanılır. Dięer durumlar için önerilmez.

3.1.3. Ortalama atama (mean imputation, ortalama deęerlendirme)

Eksik verileri işlemek için bir başka basit yöntem, bir deęişken için eksik deęerlerin basitçe deęişken için aritmetik ortalama ile deęiştirildięi ortalama atamadır. Deęişkenin koşulsuz ortalamasını uygulamak, gözlemlenen tüm verileri kullanırken hala geleneksel tam veri analizi yöntemlerini kullanarak verileri analiz etmeyi mümkün kılar. Hem liste boyunca silme hem çift silmede olduęu gibi, ortalama atama, bir MCAR varsayımı altındaki ortalamanın yansız tahminlerini verecektir. Bununla birlikte, ortalama atama, aynı deęer, tüm eksik deęerler için geçerli olduęu için verilerin deęişkenliğini azaltır. Bu, varyans için aşağı doęru bir eğilim yaratır ve aynı zamanda kümedeki dięer deęişkenlerle kovaryansları da olumsuz etkiler. Bu nedenle, ortalama atama, Ortalama hariç deęişkenlik nedeniyle çok düşük olan standart hataların yanı sıra hemen hemen tüm dięer parametreler için yanlış tahminler verecektir (Buuren, 2012).

Bu nedenle, ortalama atama, mevcut verilerden daha fazla yararlanırken, dięer iki teorik deęerlendirme kriterinden hiçbirini etkilememektedir. Bununla birlikte, kullanımı kolaydır ve karmaşık hesaplamalar gerektirmez. Sınırlı istatistik bilgisine sahip olan araştırmacılar, ortalama atamayı kullanmayı daha çok tercih etmektedirler

Demir (2013), basit atamaya dayalı yöntemler içerisinde en bilindik olanının ortalama atama yöntemi olduęunu vurgulamıştır.

Ortalama atama yönteminde, her bir deęişken için veri setindeki eksiksiz verilerin ortalaması, bu deęişkene ilişkin eksik verilerin yerine atanmakta ve böylece eksiksiz bir

veri setine ulařılmaktadır. Bu durumda, her bir deęişken için, bu şekilde yapılan atama sonucunda eksiksiz gözlemlerin ortalaması ile tamamlanmış veri setinden elde edilen ortalama birbirine eşit olmaktadır. İlgili deęişkende eksik verilerin hepsine aynı deęerin (ortalamanın) atanması merkeze doęru bir yığılmaya yol açtığından, tamamlanmış veri setinden elde edilen varyans, gerçek varyanstan düşük çıkmaktadır (Little ve Rubin, 1987).

Ayrıca bu yöntem, varyansın düşmesine paralel olarak kovaryans ve korelasyon deęerlerini de düşürmekte ve tamamen rastgele eksik koşulunda bile parametre tahminlerinde yanlılığa yol açmaktadır (Akbař ve Tavřancıl, 2015). Veri setindeki eksik miktarı arttıkça yanlılık da artmaktadır (Pigott, 2001; Enders, 2010).

Ortalama atama eksik veriler için hızlı ve basit bir düzeltmedir. Ortalama atama genel olarak sınırlı sayıda veri eksik olduęunda hızlı bir düzeltme olarak kullanılmalıdır.

řimdi ortalama atamayı daha önce verilen Çizelge 3.1. ile izah edelim

Çizelge 3.5. Ortalama atama öncesi veri tabanı

Satır No	İsim	Puan	Ders
1	A	70	Matematik
2	B	55	İngilizce
3	C	-	Bilgisayar
4	D	80	-
5	E	30	Fizik
6	-	-	Matematik

Çizelge 3.5.'de öğrencilerin puan ortalamalarını bulmaya çalıştığımızda 3 ve 6.satırların eksik veri olduğunu görürüz. Geriye kalan verilerin ortalamalarına bakıldığında; ortalama 58,75 olarak bulunur. Puan ondalık olmayacağından ortalama 59' a yuvarlanır ve eksik veriler yerine 59 yazılır ve Çizelge 3.6.'daki gibi bir durum ortaya çıkar.

Çizelge 3.6.Ortalama atama sonrası veri tabanı

Satır No	İsim	Puan	Ders
1	A	70	Matematik
2	B	55	İngilizce
3	C	59	Bilgisayar
4	D	80	-
5	E	30	Fizik
6	-	59	Matematik

3.1.4. Regresyon atama (regression imputation)

Yukarıda tartışılan eksik verilerin ele alınması için kullanılan geleneksel yöntemlerin hepsinin ortak bir özelliği vardır ki, herhangi bir karmaşık hesaplama veya verilerin manipülasyonlarını gerektirmezler ve veri analizinde kolayca uygulanırlar. Temel olarak, istatistiksel olarak zayıf olmakla beraber kullanımları kolay ve anlaşılabilirlerdir. Bununla birlikte, geleneksel yöntemlerden olan ortalama atama, eksik değerler yerine bir anlamda makul değerler atayarak, veri setinin doğru yönde temsil edilmesinde önemli bir adım atmaktadır. Bu nedenle birçok araştırmacı tarafından tercih edilmektedir (Buuren, 2012).

Regresyon ile atama yönteminde veri setindeki eksik veriler, eksik veri içermeyen diğer veriler kullanılarak kurulan bir regresyon denklemine göre bulunmaktadır. Regresyon denklemi, eksik veri içeren değişkenin tamamlanan, eksik veri içermeyen diğer değişkenlerin ise tamamlayıcı konumunda olacağı şekilde kurulur. Tamamlanan değişkende eksik olan gözlemler, diğer değişkenlere ait değerlerin bu denklemde yerine konulmasıyla tahmin edilirler.

Ortalama atamayı geliştirmenin bir yolu, diğer değerlerin gözlenen değerleri göz önüne alındığında değişken için koşullu ortalamayı uygulamaktır. Bu yaklaşım genel olarak regresyon atama (regresyon imputation) olarak bilinir. Çünkü eksik verileri, eksik olmayan verilerin regresyonundaki denklemini kullanarak tamamlar (Schafer ve Graham 2002; Allison, 2009). Doğrusal regresyon, bağımlı değişkenin bağımsız değişkenlerce doğrusal bir eşitlikle açıklanması yöntemidir. Geleneksel eksik veri yöntemlerinden farklı

olarak, regresyon atama, MAR'daki eksikliği etkileyen değişkenlerin, kullanılan istatistiksel modellere dâhil edilmesi koşuluyla, hem MCAR hem de MAR altında yansız parametre tahminleri verir. Aynı zamanda eksik veriye yönelik atamaları iyileştirmek için mevcut verilerin tümünü kullanarak mevcut verilerin kullanımını en üst düzeye çıkarır. Regresyon tahmininin dezavantajı ise, eksik değerler için koşullu araçların kullanılmasından dolayı, tüm değerler regresyon çizgisi etrafında olacak ve bu durumda değişkenler arasındaki ilişkileri doğal olarak güçlendirecektir. Bu, bir regresyon denkleminde standart hataların küçümsenmesine (daha az değer vermeyi) yanı sıra, verilerin değişkenliğinin hafife alınmasına yol açacaktır (Lee ve ark., 1994; Buuren,2012).

Stokastik regresyon ile atama (stochastic regression imputation):

Bu yöntemde regresyon ile atamada bulunan doğrusal denkleme normal dağılım gösteren bir hata terimi eklenmektedir. Böylece eksik verilerin regresyonla atanmasında karşılaşılan hata varyansının sıfır olması sorunu da ortadan kalkmaktadır. Eklenen bu hata terimi sadece veri setindeki varyansı artırmakla kalmayıp, aynı zamanda yanlılığı da düşürmektedir (Enders, 2010).

Baraldi ve Enders (2010), stokastik regresyonla değer atama tekniğinin özellikle MCAR ve MAR koşullarında, regresyonla atama tekniğine göre daha iyi sonuçlar vereceğini ve bu yöntemin yansız tahminlerde bulunabileceğini belirtmişlerdir.

Regresyon ataması, ortalama atamaya göre ileri bir adımdır. Hem ortalama atamaya nispeten kullanıcı dostu kalır hem de karmaşık hesaplamalar veya hesaplama açısından yoğun algoritmalar gerektirmez. Bununla birlikte, standart hatalarla ilgili problem regresyondaki gibi devam etmektedir.

Regresyon atamasını verilerle açıklamak için Çizelge 3.7. verilmiştir.

Çizelge 3.7.Regresyon atama öncesi veri tabanı

Boy	Kilo
151	63
174	81
138	56

Çizelge 3.7.Regresyon atama öncesi veri tabanı(devamı)

Boy	Kilo
186	91
128	47
136	57
179	76
152	72
131	48
163	100
153	-
-	83

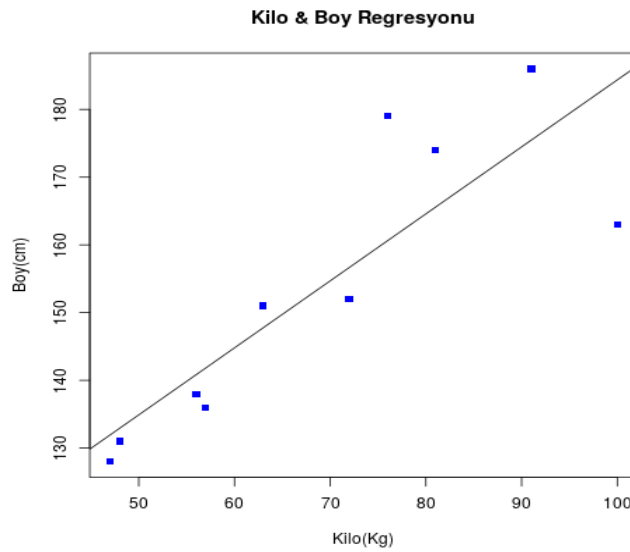
Çizelge 3.7.'de belli sayıda kişinin boy kilo değerleri verilmiştir. Boy ve kiloda bilinen değerler regresyon analizine tabi tutulduklarında;

$Kilo = -44,038 + 0,736 * Boy$ şeklinde bir regresyon denklemi ve

Şekil 3.1.'deki kilo-boy grafiği ortaya çıkmaktadır. Bu tablodaki değerler regresyon atamaya tabi tutulduklarında;

$Kilo = -44,038 + 0,736 * Boy$

Çizelge 3.7.'deki eksik değerler bu denklem yardımıyla tamamlanarak Çizelge 3.8. oluşturulur.



Şekil 3. 1. Boy kilo regresyon grafiği.

Çizelge 3.8. Regresyon atama sonrası veri tabanı

Boy	Kilo
151	63
174	81
138	56
186	91
128	47
136	57
179	76
152	72
131	48
163	100
153	68
174	83

3.1.5. Tahmini ortalama eşleştirme (sıcak deste-hot deck imputation)

Eksik verileri geleneksel yöntemlerden daha uygun bir şekilde ele almanın başka bir yolu da Tahmini Ortalama Eşleştirme (Hot Deck Imputation)'dir. Tahmini ortalama eşleştirme, verileri birbirine benzeyen bireylerden eksik veriye sahip bireyin eksik verilerinin yerine verileri tam olan benzer bireyin verilerinin yazılması durumudur. Pratikte, tahmini ortalama eşleştirme (hot deck imputation), eksik değere sahip olan bireye benzer bir dizi birey toplamak için her eksik değer için verileri tarar. Daha sonra, bu adaylardan birinin gözlemlenen değeri, eksik bir veri yerine tam bir veri seti yaratacak şekilde belirler (Buuren, 2012).

Regresyon atama yöntemlerine göre tahmini ortalama eşleşmenin önemli bir avantajı, eksik veri yerine kullanılan değerlerin mevcut verilerden alınmasıdır. Tahmini ortalama eşleştirme, modelin örtülü olmasından ve bu nedenle de imputasyonların üretildiği açıkça belirtilmiş bir modeli gerektirmediğinden hatalı tanımlama ve etkileşim etkilerine karşı daha sağlamdır. Ayrıca tahmini ortalama eşleştirme yaklaşık olarak tarafsız parametreler ürettiği, mevcut tüm verileri kullandığı ve veri MCAR veya MAR

olduğunda deęişkenlik üzerindeki minimize etme etkisine sahip olmadığı için eksik veri yöntemlerine ait üç teorik kriteri sağlamaktadır (Buuren, 2012).

Tek deęişkenli eksik veriler için, yöntem aynı zamanda hesaplama açısından hızlı ve kullanımı nispeten kolaydır. Çok deęişkenli eksik veriler için, yöntem daha karmaşıktır.

Little ve Rubin (1987) sıcak deste kavramını birkaç veri yükleme yöntemini ifade etmek için kullanmışlardır. Deste kavramı veriyi saklamak için kullanılan delikli bilgisayar kartlarından gelmektedir. Eksik verilerin doldurulması için hâlihazırda kullanılan desteden yeni deęerlerin seçilmesi sıcak deste yöntemlerini oluşturur. Bunlardan ilki eksik veriyi gözlenen verilerden rastgele seçilen bir tanesi ile doldurmaktır. Benzer şekilde dięer bir yöntem ise eksik verileri, deęişkenleri ilişkilendirerek seçilecek rastgele deęerin uygun veriden seçilmesidir. Örneğin yapılan bir çalışmaya katılan kadınlardan bazısının yaş bilgileri eksik ise bu veri setindeki eksik yaş bilgilerini tamamlamak için yaş bilgileri elde edilen kadınların yaş bilgileri arasından rastgele seçim yapılır. Üçüncü yöntem ise en yakın komşu algoritmasının kullanımınıdır. Bu yöntemle ilgili bilinen problemlerden biri standart hataların eksik tahmin edilmesidir. Bunun nedeni yüklenen verilerin hâlihazırda gözlenen verilerden seçilmesinden kaynaklı deęişkenliğin azalması olarak gösterilebilir. Ayrıca veri seti MCAR olmadığında yapılacak olan veri yüklemeleri önceden kestirilemeyecek yanlılığa sebep olur.

Tahmini ortalama eşleştirmede eksik veriler için elde bulunan mevcut veriler kullanılarak uygun atama yapılmaktadır. Veri seti içerisinden, eksik verilerin bulunduğu yerlere, o veriye benzeyen verilerin konması şeklindedir. Böylece veri matrisindeki eksik veriler benzer verilerle tamamlanır.

Bal (2003), tahmini ortalama eşleştirme yönteminin liste boyunca silme, çift silme ve ortalama atama yöntemlerinden daha üstün olduğunu vurgularken, avantajlarını da deęişkenlerin ölçüm düzeylerinin korunması, kavramsal basitlik ve tamamlanmış veri matrisinin elde edilmesi olarak saymaktadır. Ayrıca tamamlanmış veri matrisi sayesinde standart istatistiksel analizlerin uygulanabilirliğini vurgulamıştır. En önemli dezavantajında ‘benzerlik’ kavramının tanımlanmasındaki güçlük olduğunu açıklamıştır.

3.1.6. Soğuk deste ataması (cold deck imputation)

Mevcut veri setindeki eksik verileri gidermek için elde bulunan başka bir veri seti yardımıyla bu eksik veriler tamamlanabilir. Rubin (1976) başka bir veri setinden elde edilen değerleri kullanarak mevcut veri setindeki eksik verileri tamamlamayı soğuk deste olarak adlandırmıştır. Bu duruma en uygun senaryo anket çalışmalarıdır. Bunun dışında bazı çalışmalarda araştırmacılar veri setini araştırma ve doğrulama aşamalarında kullanmak için bölerler. Böylelikle araştırma sonucunda bulunan modelin doğruluğunu araştırmak için kullanılan veri seti soğuk deste olarak adlandırılır.

3.1.7. En yakın komşu yöntemi (nearest neighbor)

Tahmini ortalama eşleştirme yöntemine benzerdir. Hatta bazı kaynaklarda tahmini ortalama eşleştirme yönteminin farklı bir uygulaması olarak verilmektedir. En yakın komşu yöntemi, aslında bir sınıflandırma yöntemidir. En genel haliyle gözlemler arasındaki birim uzaklıklar, kovaryans değerleri üzerinden elde edilir. Daha sonra eksik veri değeri, en yakın gözlemden elde edilmiş değer ile tamamlanır. Birimler arasındaki uzaklıklar için maksimum sapma, mahalnobis gibi uzaklık yöntemleri kullanılabilir (Little ve Rubin, 2002). Böylece katkıda bulunan değişkenlerin göreceli önem düzeyleri uygun bir şekilde yansıtılmış olur. Elde edilmiş gözlemlerin tam kayıtlı olmasına gerek yoktur. Bazı değişkenler, yerine koymada ya da farkın hesaplanmasında göz ardı edilebilir veya bunların katkıları çok küçük olabilir (Satici, 2009).

3.1.8. Son gözlemi ileri taşıma (last observation carried forward)

Son gözlemi ileri taşıma yöntemi özellikle uzun süreli çalışmalarda karşılaşılabilecek bir yöntemdir. İsminden de anlaşılacağı gibi tekrarlı gözlemlerde eksik olan değere bir önceki değeri atama ile veriyi tamamlar. Genellikle sağlık çalışmalarında ve klinik denemelerinde kullanılmaktadır. Bu yöntem, son elde edilen değeri ileriye taşıyarak belli aralıklardaki süre içerisinde durumun hiç değişmediğini varsaymaktadır. Ancak deneysel çalışmalar bu durumun geçerli olmadığını göstermiştir (Mallinckrodt ve ark., 2001; Liu ve Gould, 2002; Cook ve ark. 2004). Özellikle eksik veri oranının fazla olması çalışmayı olumsuz yönde etkileyecektir (Enders, 2010). Son gözlemi ileri taşıma

yönteminin tek getirisi uzun zamanlı çalışmalardaki eksik gözlemleri elde edebilmenin en kolay yolu olmasıdır.

3.2. Modern Atama Yöntemleri

Modern atama yöntemleri geleneksel yöntemlerden çok sonraları çıkmış çoklu atama (Multiple Imputation) ve beklenti maksimizasyonu (EM) yöntemlerinden oluşmaktadır. Bu yöntemler model tabanlı olup her ikisinde de birden fazla işlem vardır.

3.2.1. Çoklu atama (multiple imputation)

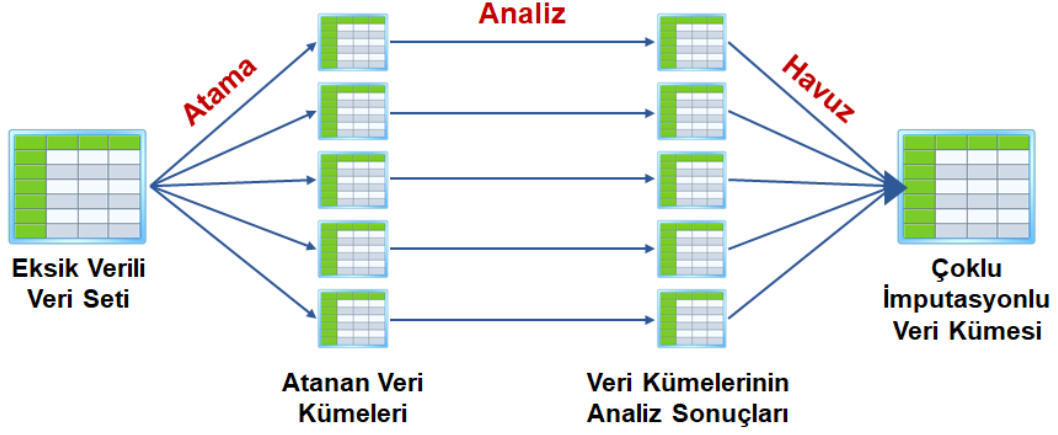
Rubin (1987)'in geliştirdiği çoklu atama yönteminde veri setindeki eksik verilerin yerine birden fazla değer atanması söz konusudur. Ayrıca çoklu atama yönteminde, verinin eksik olmasından kaynaklanan belirsizliğe bağlı olarak tek ve kesin bir değer atanması yapılmamaktadır.

Çoklu atama, $m > 1$ sayıda eksiksiz veri seti elde edilecek şekilde atama işleminin yapılması, elde edilen m farklı veri setinin standart yöntemlerle analiz edilmesi ve elde edilen sonuçların birleştirilmesinden oluşan üç aşamalı bir süreçtir (Schafer ve Graham, 2002).

Çoklu atamada regresyonla atama işlemi birden fazla kez tekrarlanarak her bir veri kümesi daha sonra kullanılmak üzere saklanır. Elde edilen tüm verilerin hata miktarları hesaplanır. Ardından bu verilerin ortalamaları alınarak gerçek veri kümesi elde edilir. Buradaki tek ayrıntı standart hatanın hesaplanması sırasında iki veri kümesinin birleştirilmesi için iki standart hata miktarı toplanıp karekökleri alınır. Kısacası çoklu atama, daha iyi sonuç elde etmek için regresyonla atamanın birden fazla yapılması olarak düşünülebilir (Şeker ve Eşmekaya, 2017).

Yukarıda tartışılan geleneksel eksik veri yöntemleriyle ilgili genel bir olumsuzluk, eksik değer yerine tek bir değer atanmasıdır. Buna çözüm olarak da Rubin (1987), her eksik veri noktası için tek bir değer uygulamak yerine, m farklı tam veri setleri oluşturmak için m farklı değerlerin uygulanması gerektiğini önermiştir. Her bir tam veri seti daha sonra standart istatistiksel yöntemler kullanılarak ayrı ayrı analiz edilir ve parametreler

için sonuçlar birleştirilir. Bu prosedür, çoklu atama (MI) olarak bilinir ve Şekil 3.2.'de olduğu gibi tarif edilir.



Şekil 3.2. Çoklu atama aşamaları.

Şekil 3.2.'de gösterildiği gibi çoklu atama 3 aşamadan oluşur.

1. **Atama:** Tekli atamaya (single imputation) benzerdir, eksik veriler atanır. Fakat atanılan veriler bir kez yerine “m” defa dağıtılır. Bu adımın sonunda “m” veri kümeleri tamamlanmalıdır.
2. **Analiz:** “m” veri kümesinin her biri analiz edilir. Bu adımın sonunda m analiz olmalıdır.
3. **Havuzlama (pooling):** ”m” sonuçları, verinin ortalama, varyans ve güvenilirlik aralığı hesaplanarak bir sonuca bağlanılır.

Çoklu atama, gözlenen değerler m farklı veri kümesi arasında aynı kaldığından, ölçülen değerlerin ve gözlenen değerlerin farklılaşmasına izin verir. Bu, her bir değerlendirme, parametreleri tahmin etmek için bir tarafsız veri noktası yaratacağı için, sürekli olmamakla beraber, eksik veri sorununu çözecektir. Böylece, m arttıkça, parametre tahmininin etkinliği artacaktır. Bu verimlilik, çoklu çarpımlarla ortaya konan değişkenlik ile birlikte doğrudan toplanan sonuçlardan doğru çıkarım yapılmasına izin verir (Rubin, 1987).

3.2.2. Beklenti maksimizasyonu (em: expectation maximization)

Eksik verileri, en çok olabilirlik kestirimleri ile dolduran iki aşamalı bir yöntem olup, ilk aşama olan beklenti aşamasında, eksik veriler beklenen değerlerle tamamlanır. İkinci aşama olan maksimizasyon aşamasında ise, ilk aşamada tahmin edilen değerler kullanılarak parametre tahmini yapılır (Enders, 2010).

Beklenti maksimizasyon yönteminde, eksik değerlerin gözlenen değerler üzerinden kurulan bir regresyon denklemiyle tahmin edildiği beklenti adımı ve bu sayede ulaşılan eksiksiz veri seti üzerinden regresyon denkleminin yeniden kurulduğu maksimizasyon adımı tekrarlı bir süreçte birbirini izlemektedir (Enders, 2010).

EM algoritmasının çalışması için öncelikle verilere ait ortalama vektörü ve kovaryans matrisinin başlangıç değerlerinin var olması gerekmektedir. Bu nedenle, öncelikle eksik veri silme yöntemleri yardımıyla bu iki parametre elde edilir. Bu iki parametre yardımıyla, var olan veriler üzerinden eksik verileri tahmin etmek amacıyla regresyon denklemi kurulur. Daha sonra regresyon denklemi yardımıyla eksik veriler için tahmini değerler bulunur. Böylece tam bir veri seti elde edilir ve E adımı tamamlanmış olur. M adımında ise elde edilen tam veri seti üzerinden ortalama vektörü ve kovaryans matrisi için en çok olabilirlik tahminleri elde edilir ve bu iki değer başlangıçta kullanılan değerlerin yerine konarak süreç tekrarlanır. Bu süreç, son adımda elde edilen ortalama vektörü ve kovaryans matrisi değerleri ile bir önceki adımda elde edilen değerler arasında anlamlı bir fark kalmayınca kadar tekrarlanır. Her tekrarda, bir önceki tekrardan elde edilen bilgi kullanıldığından daha fazla bilgi içermektedir.

EM algoritmalarının en önemli dezavantajı ise standart hata tahminini yapmamasıdır (Dempster, Laird ve Rubin, 1977; Allison, 2009, 2012).



4. MATERYAL ve YÖNTEM

4.1. Veri Üretimi

Çalışmada türetilmiş verilerden istifade edilmiştir. Veri üretimi belli şartlar dâhilinde R Language paket programları kullanılarak yapılmıştır.

Veri setleri; birim sayıları sırasıyla 50, 100, 250, 500 ve 1000 birim içerecek şekilde, normal dağılım, üstel dağılım, gamma dağılımı ve multinominal dağılım varsayımlarından yararlanılarak türetilmiştir.

Normal dağılımda *norm* komutu ile dört değişkenli bir veri seti üretilmiştir. Bu dağılımda ilgili komutlarla ortalama, standart sapma, korelasyon katsayısı, kovaryans ve regresyon belirlilik katsayısı hesaplanmıştır.

Üstel dağılımda *rexp* komutu ile dört değişkenli bir veri seti üretilmiştir. Bu dağılımda ilgili komutlarla oran(ters ölçek), korelasyon katsayısı, kovaryans ve regresyon belirlilik katsayısı hesaplanmıştır.

Gamma dağılımında *rgamma* komutu ile dört dört değişkenli bir veri seti üretilmiştir. Bu dağılımda ilgili komutlarla ölçek, korelasyon katsayısı, kovaryans ve regresyon belirlilik katsayısı hesaplanmıştır.

Multinominal Dağılımda ise *rmultinom* komutu ile dört değişkenli bir veri seti üretilmiştir. Bu dağılımda ortalama, standart sapma, ilişki(korelasyon) katsayısı ve, kovaryansları hesaplanmıştır.

4.2. Veri Eksiltilmesi

N= 50, 100, 250, 500, 1000 olacak şekilde rastgele türetilen her bir dağılıma özel tam veri setlerinin en sonuncusundaki veriler R Language paket programları kullanılarak eksik veri mekanizmalarına (MCAR, MAR, MNAR) göre %1, %5 ve %10 oranlarında eksiltilmiş ve eksik veri analizine esas olacak şekilde eksik veri setleri elde edilmiştir. Eksiltelen bu veri setlerinin de her dağılıma, eksik veri mekanizmasına, eksik veri miktarına özel parametreleri hesaplanmıştır.

4.3. Veri Tamamlanması

Eksiltelen veri setlerine sırasıyla liste boyunca silme, ortalama atama, regresyonla atama, tahmini ortalama eşleştirme ve çoklu atama yöntemleri uygulanmış ve tamamlanan bu veri setlerinin de parametreleri hesaplanmıştır. Multinomial dağılım kesikli dağılım olduğundan regresyon ile atama uygulanmamıştır. Çift silme yöntemi eksik veri seti bir olarak çalışıldığından liste boyunca silme ile aynı sonucu vereceğinden üzerinde çalışılmamıştır.

4.4. Parametrelerin Hesaplanması

Ortalama ve standart sapma hesaplanırken R'ın base paketinde yer alan *mean* ve *sd* komutları kullanılmıştır.

Korelasyon ve kovaryans ayrı ayrı hesaplanırken sırasıyla *cor* ve *cov* komutları kullanılmıştır. Korelasyon katsayısı ile kovaryansın hesaplanmasında gerçek değer kendisi ile korelasyonu ya da kovaryansı gerçek değer olarak kabul edilmiş, gerçek değer ile her atama yönteminin sonucunda çıkan değerlerle olan korelasyonları ya da kovaryansları bulunmuştur.

Belirlilik katsayısının (Multiple R²) hesaplanmasında ise önce regresyon *lm* komutuyla kurulmuştur. Belirlilik katsayısının (Multiple R²) hesaplanmasında dördüncü veri seti bağımlı değişken olarak düşünülmüş, bu veri setinin gerçek değeri ve eksik veri yönetme yöntemleri uygulandıktan sonraki değerlerinin diğer üç veri seti ile olan çoklu regresyonu hesaplanmış ve çıkan değerlerin yakınsaklık durumları incelenmiştir.

Üstel dağılımda oran (ters ölçek) hesaplaması yapılırken ortalama=1/oran formülünden yararlanılmış, önce ortalama bulunmuş akabinde tersi elde edilmiştir.

Gamma dağılımında ise, ölçek bulunurken önce ortalamalar bulunmuş akabinde gamma dağılımında işleme sokulan k şekil parametresinden dolayı bu değer k'ya bölünmüştür.

Hesaplanan tüm parametreler birbirleriyle karşılaştırılmış ve eksik veri analizi yöntemlerinin etkinliği yakınsama durumuna göre belirlenmiştir.

5. BULGULAR

5.1. Normal Dağılım

Normal dağılıma göre rastgele üretilen, eksik veri mekanizmalarına göre eksiltelen ve eksik veri analiz yöntemleri ile tamamlanan verilerin eksik veri mekanizmasına ve eksiltileme yüzdelere göre parametrelerinde oluşan durum ile eksik veri analizi yöntemlerinin parametreler üzerindeki etkinliklerinin yakınsama durumları her eksik veri mekanizmasına özel olacak şekilde sırasıyla Çizelge 5.1., Çizelge 5.2. ve Çizelge 5.3.'de gösterilmiş ve çizelgeler ayrı ayrı yorumlanmıştır.

Çizelge 5.1. Normal dağılımda MCAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=50	1%	A.O							
		S.S.							
		Kor. Kov. R ²		Herhangi Bir Eksik Veri Bulunmadığından İşlem Yapılmamıştır.					
	5%	A.O	10.2180	10.2728	10.2728	10.2728	10.2683	10.2591*	10.2642
		S.S.	0.9512	0.9183	0.9183	0.8994	0.9007	0.9362*	0.9011
		Kor.	1.0000	**	**	0.9455	0.9435	0.8869	0.9507*
		Kov. R ²	0.9048 0.0720	** **	0.8434* 0.0880	0.8090 0.0820*	0.8085 0.0872	0.7899 0.1108	0.8150 0.0874
	10%	A.O	10.2180	10.2600	10.2600	10.2600	10.2371*	10.2845	10.2617
		S.S.	0.9512	0.8999	0.8999*	0.8527	0.8569	0.8592	0.8585
		Kor.	1.0000	**	**	0.8964*	0.8814	0.8906	0.8839
		Kov. R ²	0.9048 0.0720	** **	0.8098* 0.1101	0.7271 0.1012	0.7185 0.1188	0.7279 0.0087*	0.7219 0.1027
	N=100	1%	A.O	10.0108	10.0157	10.0157	10.0157	10.0200	10.0140*
S.S.			1.0063	1.0102	1.0102	1.0051	1.0060*	1.0053	1.0086
Kor.			1.0000	**	**	0.9988	0.9958	0.9995*	0.9912
Kov. R ²			1.0127 0.0549	** **	1.0206 0.0607	1.0103 0.0589	1.0082 0.0624	1.0112* 0.0575*	1.0062 0.0656
5%		A.O	10.0108	9.9993	9.9993	9.9993	9.9947	10.0474	10.0079*
		S.S.	1.0063	1.0194	1.0194	0.9934	0.9959	1.0493	1.0000*
		Kor.	1.0000	**	**	0.9871*	0.9813	0.9114	0.9762
		Kov. R ²	1.0127 0.0549	** **	1.0393* 0.0670	0.9868 0.0623	0.9835 0.0716	0.9625 0.0480*	0.9825 0.0698
10%		A.O	10.0108	10.0319	10.0319*	10.0319*	10.086	10.0665	10.0488
		S.S.	1.0063	1.0061	1.0061*	0.9540	0.9567	0.9936	0.9751
		Kor.	1.0000	**	**	0.9479*	0.9445	0.9311	0.8914
		Kov. R ²	1.0127 0.0549	** **	1.0124* 0.0643	0.9101 0.0592	0.9094 0.0696	0.9311 0.0499*	0.8748 0.0731

Çizelge 5.1. Normal dağılımda MCAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri (devamı)

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=250	1%	A.O	9.9722	9.9731	9.9731*	9.9731*	9.9732	9.9709	9.9771
		S.S.	1.0336	1.0299	1.0299*	1.0258	1.0259	1.0266	1.0268
		Kor.	1.0000	**	**	0.9924	0.9927	0.9957*	0.9919
		Kov.	1.0685	**	1.0608*	1.0523	1.527	1.0567	1.0528
		R ²	0.0091	**	0.0085	0.0085	0.0086	0.0087	0.0088*
	5%	A.O	9.9722	9.9612	9.9612	9.9612	9.9602	9.9433	9.9640*
		S.S.	1.0336	1.0248	1.0248	0.9998	0.9999	1.0279*	1.0017
		Kor.	1.0000	**	**	0.9672	0.9682	0.9591	0.9697*
		Kov.	1.0685	**	1.0502*	0.9996	1.0008	1.0192	1.0041
		R ²	0.0091	**	0.0070	0.0067	0.0073*	0.0108	0.0073*
	10%	A.O	9.9722	9.9537	9.9537	9.9537	9.9515	9.9689*	9.9476
		S.S.	1.0336	1.0376	1.0376*	0.9841	0.9844	1.0267	0.9921
Kor.		1.0000	**	**	0.9520	0.9529*	0.8834	0.9452	
R ²		0.0091	**	0.0067	0.0062	0.0073*	0.0056	0.0070	
N=500	1%	A.O	9.9859	9.9825	9.9825	9.9825	9.9825	9.9783	9.9839*
		S.S.	0.9526	0.9491	0.9491*	0.9444	0.9444	0.9459	0.9450
		Kor.	1.0000	**	**	0.9913	0.9913	0.9915	0.9946*
		Kov.	0.9075	**	0.9009*	0.8919	0.8919	0.8935	0.8954
		R ²	0.0015	**	0.0013	0.0012	0.0014*	0.0013	0.0014*
	5%	A.O	9.9859	9.9864	9.9864*	9.9864*	9.9867	9.9816	9.9927
		S.S.	0.9526	0.9573	0.9573*	0.9330	0.9330	0.9688	0.9398
		Kor.	1.0000	**	**	0.9793	0.9794*	0.9101	0.9743
		Kov.	0.9075	**	0.9164*	0.8705	0.8705	0.8400	0.8724
		R ²	0.0015	**	0.0009	0.0008	0.0010	0.0011	0.0012*
	10%	A.O	9.9859	9.9813	9.9813	9.9813	9.9814	9.9945	9.9876*
		S.S.	0.9526	0.9616	0.9616*	0.9122	0.9123	0.9648	0.9217
Kor.		1.0000	**	**	0.9575*	0.9571	0.9037	0.9463	
Kov.		0.9075	**	0.9247*	0.8321	0.8319	0.8307	0.8310	
R ²	0.0015	**	0.0015*	0.0014	0.0017	0.0014	0.0015*		
N=1000	1%	A.O	10.0605	10.063	10.0630	10.0630	10.0629	10.0672	10.0611*
		S.S.	0.9957	0.9967	0.9967	0.9917	0.9918	0.9955*	0.9924
		Kor.	1.0000	**	**	0.9959*	0.9958	0.9935	0.9945
		Kov.	0.9915	**	0.9935*	0.9836	0.9834	0.9849	0.9829
		R ²	0.0040	**	0.0043	0.0043	0.0044	0.0042*	0.0045
	5%	A.O	10.0605	10.0712	10.0712	10.0712	10.0709	10.0728	10.0710*
		S.S.	0.9957	0.9903	0.9903	0.9652	0.9653	0.9859*	0.9699
		Kor.	1.0000	**	**	0.9693*	0.9686	0.9610	0.9602
		Kov.	0.9915	**	0.9807*	0.9316	0.9311	0.9435	0.9275
		R ²	0.0040	**	0.0054	0.0052	0.0056	0.0043*	0.0060
	10%	A.O	10.0605	10.0700	10.0700	10.0700	10.0701	10.0715	10.0668*
		S.S.	0.9957	0.9968	0.9968*	0.9456	0.9458	0.9913	0.9548
Kor.		1.0000	**	**	0.9496*	0.9493	0.8996	0.9449	
Kov.		0.9915	**	0.9936*	0.8941	0.8941	0.8880	0.8984	
R ²	0.0040	**	0.0049	0.0044*	0.0055	0.0022	0.0055		

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.1. incelendiğinde, aritmetik ortalama yakınsamasında, tüm yöntemlerin gerçek değerle farklılaşmasının anlamlı olacak şekilde fazla olmadığı, ancak sırasıyla çoklu atama ve tahmini ortalama eşleştirme yöntemlerinin diğer yöntemlerden daha başarılı olduğu görülmektedir. Özellikle veri miktarı arttıkça çoklu atamanın etkinliğinin yüksek olduğu gözlemlenmiştir.

Standart sapma yakınsamasında da tüm yöntemlerin gerçek değerle farklılaşmasının anlamlı olacak şekilde fazla olmadığı liste silme ile tahmini ortalama eşleştirme yöntemleri diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında da tüm yöntemlerin gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla ortalama atama ve çoklu atamanın diğer yöntemlere göre daha etkin oldukları, ortalama atamada veri sayısının artmasıyla etkinliğin gerçek değere yakınsamasının daha da belirgin hale geldiği gözlemlenmiştir.

Kovaryansların karşılaştırılmasında da gerçek değer ile yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, ancak liste silme yönteminin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Belirlilik katsayılarının (Multiple R²) karşılaştırılmasında ise gerçek değer tüm yöntemlerle karşılaştırılmış, tahmini ortalama eşleştirme yönteminin büyük veri setlerinin yüksek oranlı eksik veri olduğu durumda düşük de olsa önemli bir farkın olduğu gözlemlenmiştir. Ayrıca, sırasıyla çoklu atama ve regresyon ile atamanın diğer yöntemlerden daha etkin oldukları gözlemlenmiştir.

Çizelge 5.2. Normal dağılımda MAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=50	1%	A.O							
		S.S.							
		Kov. R ²		Herhangi Bir Eksik Veri Bulunmadığından İşlem Yapılmamıştır.					
	5%	A.O	10.2180	10.2062	10.2062	10.2062	10.2032	10.2151*	10.1988
		S.S.	0.9512	0.9680	0.9680	0.9480	0.9482	0.9491*	0.9487
		Kor.	1.0000	**	**	0.9966	0.9958	0.9978*	0.9934
		Kov.	0.9048	**	0.9370	0.8988	0.8982	0.9009*	0.8966
		R ²	0.0720	**	0.0738	0.0735	0.0741	0.0718*	0.0746
		A.O	10.2180	10.1531	10.1531	10.1531	10.1463	10.1239	10.1725*
	10%	S.S.	0.9512	0.9586	0.9586*	0.9084	0.9114	0.9646	0.9260
		Kor.	1.0000	**	**	0.9549*	0.9381	0.8731	0.9382
		Kov.	0.9048	**	0.9190*	0.8252	0.8133	0.8011	0.8265
R ²		0.0720	**	0.1099	0.1039*	0.1157	0.1221	0.1232	
A.O		10.0108	10.0051	10.0051	10.0051	10.0040	10.0124*	10.0081	
S.S.		1.0063	1.0098	1.0098	1.0047	1.0048	1.0073*	1.0053*	
N=100	1%	Kor.	1.0000	**	**	0.9983	0.9977	0.9998*	0.9996
		Kov.	1.0127	**	1.0198	1.0095	1.0089	1.0136*	1.0112
		R ²	0.0549	**	0.0564	0.0562	0.0565	0.0545*	0.0556

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.2. Normal dağılımda MAR mekanizmasına göre ve eksik veri yüzdesine göre parametre değerleri (devamı)

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama	
N=100	5%	A.O	10.0108	10.0000	10.0000	10.0000	10.0057*	10.0227	10.0191	
		S.S.	1.0063	1.0291	1.0291	1.0028	1.0042	1.0132	1.0078*	
		Kor.	1.0000	**	**	0.9964	0.9938	0.9912	0.9974*	
		Kov.	1.0127	**	1.0592	1.0057	1.0044	1.0107	1.0116*	
	R ²	0.0549	**	0.0599	0.0574*	0.0625	0.0598	0.0595		
	10%	A.O	10.0108	9.9770	9.9770	9.9770	9.9825	9.9924	10.0031*	
		S.S.	1.0063	1.0221	1.0221	0.9691	0.9702	1.0121	0.9850*	
		Kor.	1.0000	**	**	0.9630	0.9634	0.8971	0.9680*	
		Kov.	1.0127	**	1.0448*	0.9392	0.9408	0.9138	0.9597	
	R ²	0.0549	**	0.0576	0.0556	0.0597	0.0578	0.0544*		
	N=250	1%	A.O	9.9722	9.9702	9.9702	9.9702	9.9703	9.9715*	9.9741
			S.S.	1.0336	1.0371	1.0371	1.0329	1.0330	1.0359	1.0338*
Kor.			1.0000	**	**	0.9992	0.9993*	0.9943	0.9992	
Kov.			1.0685	**	1.0755	1.0669	1.0670	1.0647	1.0679*	
R ²		0.0091	**	0.0075	0.0089	0.0090*	0.0083	0.0089		
5%		A.O	9.9722	9.9501	9.9501	9.9501	9.9515	9.9428	9.9553*	
		S.S.	1.0336	1.0337	1.0337*	1.0085	1.0086	1.0229	1.0106	
		Kor.	1.0000	**	**	0.9756	0.9776	0.9742	0.9806*	
		Kov.	1.0685	**	1.0686*	1.0171	1.0192	1.0301	1.0244	
R ²		0.0091	**	0.0042	0.0041	0.0044	0.0054*	0.0050		
10%		A.O	9.9722	9.9592	9.9592	9.9591	9.9638	9.9473	9.9600*	
		S.S.	1.0336	1.0335	1.0335*	0.9802	0.9806	1.0265	0.9879	
	Kor.	1.0000	**	**	0.9483*	0.9477	0.9006	0.9326		
	Kov.	1.0685	**	1.0681*	0.9608	0.9607	0.9557	0.9523		
R ²	0.0091	**	0.0078	0.0071	0.0086*	0.0098	0.0075			
N=500	1%	A.O	9.9859	9.9836	9.9836	9.9836	9.9836	9.9908	9.9850*	
		S.S.	0.9526	0.9565	0.9565	0.9517	0.9517	0.9571	0.9524*	
		Kor.	1.0000	**	**	0.9990	0.9990	0.9932	0.9993*	
		Kov.	0.9075	**	0.9150	0.9058	0.9058	0.9057	0.9067*	
	R ²	0.0015	**	0.0015*	0.0015*	0.0015*	0.0014	0.0015*		
	5%	A.O	9.9859	9.9828	9.9828	9.9828	9.9825	9.9865*	9.9886	
		S.S.	0.9526	0.9568	0.9568*	0.9326	0.9326	0.9501	0.9393	
		Kor.	1.0000	**	**	0.9789*	0.9786	0.9662	0.9746	
		Kov.	0.9075	**	0.9156	0.8697	0.8695	0.8745*	0.8722	
	R ²	0.0015	**	0.0021	0.0020*	0.0021	0.0023	0.0021		
	10%	A.O	9.9859	9.9979	9.979	9.9754*	9.9789	10.0030	9.9851	
		S.S.	0.9526	0.9647	0.9647	0.9151	0.9152	0.9576*	0.9247	
Kor.		1.0000	**	**	0.9606	0.9602	0.9004	0.9628*		
Kov.		0.9075	**	0.9308*	0.8375	0.8372	0.8214	0.8482		
R ²	0.0015	**	0.0023	0.0021	0.0025	0.0017*	0.0022			
N=1000	1%	A.O	10.0605	10.0585	10.585	10.0585	10.0588*	10.0633	10.0570	
		S.S.	0.9957	0.9988	0.9988	0.9938	0.9938	0.9970*	0.9944*	
		Kor.	1.0000	**	**	0.9980	0.9981*	0.9978	0.9973	
		Kov.	0.9915	**	0.9976	0.9876	0.9877	0.9906*	0.9875	
	R ²	0.0040	**	0.0038	0.0038	0.0039*	0.0044	0.0038		
	5%	A.O	10.0605	10.0516	10.0516	10.0516	10.0515	10.0542*	10.0517	
		S.S.	0.9957	0.9960	0.9960*	0.9707	0.9708	0.9896	0.9755	
		Kor.	1.0000	**	**	0.9748	0.9749*	0.9641	0.9680	
		Kov.	0.9915	**	0.9920*	0.9423	0.9425	0.9501	0.9403	
	R ²	0.0040	**	0.0040*	0.0038	0.0041	0.0028	0.0045		
	10%	A.O	10.0605	10.0411	10.0411	10.0411	10.0409	10.0424*	10.0377	
		S.S.	0.9957	0.9979	0.9979*	0.9467	0.9468	0.9895	0.9557	
Kor.		1.0000	**	**	0.9507	0.9510*	0.9066	0.9433		
Kov.		0.9915	**	0.9960	0.8963	0.8967	0.8933	0.8978*		
R ²	0.0040	**	0.0036	0.0033	0.0039*	0.0033	0.0037			

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.2.'de gerçek değerlerle yöntemlere ait değerlerin aritmetik ortalaması karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ancak, sırasıyla çoklu atama ile tahmini ortalama eşleştirme yöntemlerinin diğer yöntemlerden daha başarılı olduğu, özellikle eksik veri miktarı arttıkça çoklu atamanın etkinliğinin diğer yöntemlerden daha yüksek olduğu gözlemlenmiştir.

Orijinal veri seti ve verilen yöntemlerle eksik verileri tamamlanan veri setleri standart sapmaları arasında anlamlı bir farkın olmadığı, ancak çoklu atama ile liste silme yöntemlerinin diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir. Eksik veri miktarının az olduğu durumlarda çoklu atamanın başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla çoklu atama ve regresyon ile atamanın diğer yöntemlere göre daha etkin oldukları, regresyon ile atamada veri sayısı arttıkça etkinliğin gerçek değere daha da yaklaştığı gözlemlenmiştir.

Kovaryansların karşılaştırılmasında da gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, ancak liste silmenin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Belirlilik katsayılarının (Multiple R²) karşılaştırılmasında ise gerçek değer ile yöntemlerle bulunan değerler arasındaki farkın anlamlı olacak kadar farklı olmadığı gözlemlenmiştir. Ayrıca regresyon ile atama yönteminin diğer yöntemlerden daha etkin oldukları gözlemlenmiştir.

Çizelge 5.3.Normal dağılımda MNAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örnekleme Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=50	1%	A.O							
		S.S.							
		Kor.							
		R ²							
	5%	A.O	102.180	101.293	10.1293*	10.1293*	101.177	101.027	101.140
		S.S.	0.9512	0.8586	0.8586	0.8409	0.8431	0.8654*	0.8444
		Kor.	10.000	**	**	0.8839*	0.8521	0.8026	0.8401
		R ²	0.9048	**	0.7372*	0.7071	0.6834	0.6608	0.6749
	10%	A.O	102.180	100.587	100.587	100.587	10.0632*	100.434	100.452
		S.S.	0.9512	0.8048	0.8048	0.7712	0.7763	0.8030*	0.7746
		Kor.	10.000	**	**	0.8107*	0.8103	0.7392	0.7757
		R ²	0.9048	**	0.6477*	0.5948	0.5984	0.5647	0.5716
			**	0.0973	0.0861	0.1091	0.1305	0.0922*	

Çizelge 5.3. Normal dağılımda MNAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri(devamı)

Veri Seti Örnekleme	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=100	1%	A.O	100.108	99.647	99.647	99.647	99.662	99.607	9.9699*
		S.S.	10.063	0.8991	0.8991	0.8946	0.8947	0.8954	0.8961*
		Kor.	10.000	**	**	0.8889	0.8963	0.8676	0.9139*
		Kov.	10.127	**	0.8084	0.8003	0.8071	0.7819	0.8242*
	5%	R ²	0.0549	**	0.0430	0.0427	0.0433	0.0413	0.0446*
		A.O	100.108	98.913	98.913	98.913	98.947	98.997	9.9057*
		S.S.	10.063	0.8399	0.8399*	0.8184	0.8189	0.8393	0.8245
		Kor.	10.000	**	**	0.8132	0.8253	0.8343	0.8487*
	10%	Kov.	10.127	**	0.7054*	0.6698	0.6802	0.7047	0.7043
		R ²	0.0549	**	0.0370	0.0358	0.0381	0.0286	0.0414*
		A.O	100.108	98.087	98.087	98.087	98.136	97.767	9.8295*
		S.S.	10.063	0.7834	0.7834*	0.7442	0.7440	0.7832	0.7543
N=250	1%	Kor.	10.000	**	**	0.7381	0.7519	0.5739	0.7706*
		Kov.	10.127	**	0.6138*	0.5518	0.5630	0.4523	0.5850
		R ²	0.0549	**	0.0358	0.03285	0.0389*	0.0121	0.0309
		A.O	99.722	99.382	99.382	99.382	99.374	99.325	9.9408*
	5%	S.S.	10.336	0.9920	0.9920*	0.9860	0.9861	0.9912	0.9875
		Kor.	10.000	**	**	0.9539	0.9517	0.9313	0.9585*
		Kov.	10.685	**	0.9841*	0.9722	0.9701	0.9542	0.9784
		R ²	0.0091	**	0.0136	0.0134	0.0137	0.0134	0.0126*
	10%	A.O	99.722	98.525	98.525	98.525	98.509	98.478	9.8548*
		S.S.	10.336	0.9174	0.9174	0.8931	0.8935	0.9218*	0.8980
		Kor.	10.000	**	**	0.8640*	0.8599	0.8330	0.8628
		Kov.	10.685	**	0.8416*	0.7977	0.7943	0.7937	0.8010
N=500	1%	R ²	0.0091	**	0.0172	0.0164	0.0180	0.0207	0.0150*
		A.O	99.722	97.640	97.640	97.640*	97.599	97.597	97.568
		S.S.	10.336	0.8544	0.8544*	0.8104	0.8112	0.8387	0.8166
		Kor.	10.000	**	**	0.7840*	0.7738	0.7526	0.7613
	5%	Kov.	10.685	**	0.7301*	0.6568	0.6489	0.6524	0.6426
		R ²	0.0091	**	0.0233	0.0215	0.0253	0.0203*	0.0230
		A.O	99.859	99.579	99.579	99.579	99.577	99.567	9.9590*
		S.S.	0.9526	0.9151	0.9151	0.9105	0.9106	0.9156*	0.9110
	10%	Kor.	10.000	**	**	0.9558	0.9548	0.9455	0.9592*
		Kov.	0.9075	**	0.8375*	0.8291	0.8283	0.8248	0.8325
		R ²	0.0015	**	0.0032	0.0032	0.0033	0.0028*	0.0030
		A.O	99.859	98.776	98.776	98.776	98.765	98.625	9.8818*
N=1000	1%	S.S.	0.9526	0.8430	0.8430	0.8216	0.8217	0.8442*	0.8272
		Kor.	10.000	**	**	0.8624	0.8591	0.8015	0.8668*
		Kov.	0.9075	**	0.7106*	0.6750	0.6726	0.6446	0.6831
		R ²	0.0015	**	0.0059	0.0055	0.0061	0.0081	0.0051*
	5%	A.O	99.859	97.926	97.926	97.926	97.915	97.777	9.7966*
		S.S.	0.9526	0.7822	0.7822	0.7420	0.7423	0.7989*	0.7497
		Kor.	10.000	**	**	0.7789	0.7757	0.6853	0.7838*
		Kov.	0.9075	**	0.6119*	0.5506	0.5486	0.5215	0.5598
	10%	R ²	0.0015	**	0.0073	0.0071*	0.0087	0.0085	0.0075
		A.O	100.605	100.339	100.339	100.339	100.340	10.0352*	100.323
		S.S.	0.9957	0.9645	0.9645*	0.9596	0.9596	0.9635	0.9602
		Kor.	10.000	**	**	0.9637	0.9639*	0.9625	0.9582
N=1000	1%	Kov.	0.9915	**	0.9302*	0.9209	0.9211	0.9235	0.9162
		R ²	0.0040	**	0.0037	0.0037	0.0038*	0.0030	0.0038*
		A.O	100.605	99.530	99.530	99.530	99.533*	99.465	99.534
		S.S.	0.9957	0.8978	0.8978*	0.8750	0.8751	0.8969	0.8789
	5%	Kor.	10.000	**	**	0.8787	0.8793*	0.8398	0.8760
		Kov.	0.9915	**	0.8061*	0.7657	0.7662	0.7500	0.7667
		R ²	0.0040	**	0.0025	0.0024	0.0026*	0.0018	0.0020
		A.O	100.605	98.647	98.647	98.647	9.8651*	98.585	98.623
	10%	S.S.	0.9957	0.8376	0.8376	0.7946	0.7947	0.8415*	0.8022
		Kor.	10.000	**	**	0.7979	0.7990*	0.7402	0.7828
		Kov.	0.9915	**	0.7016*	0.6314	0.6323	0.6203	0.6254
		R ²	0.0040	**	0.0027	0.0025	0.0030	0.0032*	0.0030

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.3.'de gerçek değerlerle yöntemlere ait değerlerin aritmetik ortalaması karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ancak, sırasıyla çoklu atama ile regresyon atama yöntemlerinin diğer yöntemlerden daha başarılı olduğu, gözlemlenmiştir.

Orijinal veri seti ve verilen yöntemlerle eksik verileri tamamlanan veri setleri standart sapmaları arasında anlamlı bir farkın olduğu, ortalama atama ve regresyon ile atamanın özellikle eksik veri miktarının fazla olduğu durumlarda anlamlı olacak şekilde farklılaştığı gözlemlenmiştir. Ayrıca liste silme ile tahmini ortalama eşleştirme yöntemleri diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın eksik veri miktarının düşük olduğu(%1) durumlarda anlamlı olacak şekilde fazla olmadığı diğer durumlarda anlamlı olacak kadar fazla olduğu belirlenmiştir. Ayrıca çoklu atamanın diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Kovaryansların karşılaştırmasında da tüm yöntemlerin gerçek değerle farklılaşmasının anlamlı olacak şekilde fazla olduğu, liste silmenin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Belirlilik katsayılarının (Multiple R²) karşılaştırılmasında ise gerçek değer ile yöntemlerle bulunan değerler arasındaki farkın anlamlı olacak kadar farklı olduğu gözlemlenmiştir. Ayrıca, çoklu atama yönteminin diğer yöntemlerden daha etkin olduğu gözlemlenmiştir.

5.2. Üstel Dağılım

Üstel dağılıma göre rastgele üretilen, eksik veri mekanizmalarına göre eksiltelen ve eksik veri analiz yöntemleri ile tamamlanan verilerin eksik veri mekanizmasına ve eksiltileme yüzdelerine göre parametrelerinde oluşan durum ile eksik veri analizi yöntemlerinin parametreler üzerindeki etkinliklerinin yakınsama durumları her eksik veri mekanizmasına özel olacak şekilde sırasıyla Çizelge 5.4., Çizelge 5.5. ve Çizelge 5.6.'da gösterilmiş ve çizelgeler ayrı ayrı yorumlanmıştır.

Çizelge 5.4. Üstel dağılım MCAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=50	1%	Oran							
		Kor							
		R ²							
	5%	Oran	0.5659	0.5629	0.5629	0.5629	0.5646*	0.5701	0.5629
		Kor	1.0000	**	**	0.9803	0.9809	0.9749	0.9833*
		Kov	2.3668	**	2.3716*	2.2748	2.2765	2.2688	2.2820
		R ²	0.0335	**	0.0336*	0.0333	0.0339	0.0357	0.0334*
		Oran	0.5659	0.5559	0.5559	0.5559	0.5582*	0.5532	0.5337
		Kor	1.0000	**	**	0.9709	0.9710*	0.8983	0.9508
	10%	Kov	2.3668	**	2.4849*	2.2313	2.2320	2.1578	2.2430
		R ²	0.0335	**	0.0350	0.0346*	0.0355	0.0407	0.0322
		Herhangi Bir Eksik Veri Bulunmadığından İşlem Yapılmamıştır.							
N=100	1%	Oran	0.4846	0.4834	0.4834*	0.4834*	0.4830	0.4799	0.4864
		Kor	1.0000	**	**	0.9996*	0.9994	0.9948	0.9992
		R ²	0.0241	**	0.0246*	0.0248	0.0247	0.0257	0.0235*
	5%	Oran	0.4846	0.4806	0.4806	0.4806	0.4809	0.4859*	0.4785
		Kor	1.0000	**	**	0.9901	0.9914	0.9844	0.9950*
		Kov	4.0021	**	4.1327	3.9240	3.9300	3.9258	3.9757*
		R ²	0.0241	**	0.0215	0.0211	0.0219	0.2013	0.0255*
		Oran	0.4846	0.4703	0.4703*	0.4703*	0.4702	0.4657	0.4652
		Kor	1.0000	**	**	0.9835*	0.9835*	0.9548	0.9739
	10%	Kov	4.0021	**	4.3063	3.8713	3.8754*	3.8362	3.8502
		R ²	0.0241	**	0.0243*	0.0224	0.0263	0.0293	0.0244
N=250	1%	Oran	0.5411	0.5390	0.5390	0.5390	0.5391	0.5407*	0.5389
		Kor	1.0000	**	**	0.9987	0.9988	0.9999*	0.9987
		R ²	0.0074	**	0.0071	0.0071	0.0072	0.0071	0.0074*
	5%	Oran	0.5411	0.5300	0.5300*	0.5300*	0.5293	0.5291	0.5279
		Kor	1.0000	**	**	0.9908*	0.9895	0.9830	0.9822
		Kov	2.8918	**	2.9827	2.8390*	2.8359	2.8369	2.8333
		R ²	0.0074	**	0.0095	0.0091	0.0099	0.0086*	0.0105
		Oran	0.5411	0.5268	0.5268	0.5268*	0.5255	0.5235	0.5252
		Kor	1.0000	**	**	0.9753*	0.9726	0.9270	0.9683
	10%	Kov	2.8918	**	3.0578	2.7508	2.7443	2.7232	2.7545*
		R ²	0.0074	**	0.0120	0.0112*	0.0129	0.0118	0.0130
N=500	1%	Oran	0.4898	0.4874	0.4874	0.4874	0.4871	0.4880	0.4882*
		Kor	1.0000	**	**	0.9981	0.9979	0.9957	0.9987*
		R ²	0.0068	**	0.0073	0.0071*	0.0076	0.0060	0.0077
	5%	Oran	0.4898	0.4918	0.4918	0.4918	0.4915*	0.4941	0.4945
		Kor	1.0000	**	**	0.9793*	0.9784	0.9591	0.9776
		Kov	3.6892	**	3.7252*	3.5386	3.5364	3.5289	3.5548
		R ²	0.0068	**	0.0087	0.0082	0.0093	0.0079*	0.0083
		Oran	0.4898	0.4982	0.4982	0.4982	0.4977*	0.5039	0.4992
		Kor	1.0000	**	**	0.9406	0.9367	0.9102	0.9307
	10%	Kov	3.6892	**	3.6278*	3.2642	3.2534	3.2506	3.2589
		R ²	0.0068	**	0.0137	0.0124*	0.0153	0.0149	0.0124*
N=1000	1%	Oran	0.4968	0.4963	0.4963*	0.4963*	0.4962	0.4960	0.4960
		Kor	1.0000	**	**	0.9946*	0.9946*	0.9910	0.9920
		R ²	0.0013	**	0.0011	0.0011	0.0012*	0.0012*	0.0011
	5%	Oran	0.4968	0.4911	0.4911	0.4911*	0.4911*	0.4903	0.4901
		Kor	1.0000	**	**	0.9812*	0.9812	0.9409	0.9777
		Kov	3.9738	**	4.0281*	3.8265	3.8260	3.7813	3.8356
		R ²	0.0013	**	0.0013*	0.0012	0.0013*	0.0010	0.0013*
		Oran	0.4968	0.4935	0.4935	0.4935	0.4937*	0.4878	0.4935
		Kor	1.0000	**	**	0.9582*	0.9578	0.8883	0.9493
	10%	Kov	3.9738	**	4.0545	3.6486	3.6474	3.6222	3.6568*
		R ²	0.0013	**	0.0015	0.0014*	0.0017	0.0032	0.0018

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.4.'de gerçek değerlerle yöntemlere ait değerlerin oranları karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ancak, sırasıyla regresyon ile atama ve ortalama atama yöntemlerinin diğer yöntemlerden daha başarılı olduğu, gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla ortalama atama ile çoklu atama yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir.

Kovaryansların karşılaştırılmasında da tüm yöntemlerin gerçek değerle farkının anlamlı olacak şekilde fazla olmadığı gözlemlenmiştir. Ayrıca çoklu atama ve liste silme yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir.

Belirlilik katsayısı (Multiple R²) karşılaştırılmasında ise tahmini ortalama eşleştirme dışındaki yöntemlerin yakınsamasının gerçek değerle farkının anlamlı olacak şekilde olmadığı gözlemlenmiştir. Tahmini ortalama eşleştirme yönteminde ise yüksek oranlı eksik veri setlerinde (%10) farklılığının belirginleştiği saptanmıştır. Çoklu atama yönteminin diğer yöntemlerden daha etkin olduğu gözlemlenmiştir.

Çizelge 5.5. Üstel dağılım MAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama	
N=50	1%	Oran								
		Kor		Herhangi Bir Eksik Veri Bulunmadığından İşlem Yapılmamıştır.						
		Kov								
	5%	R ²								
		Oran	0.4570	0.4581	0.4581*	0.4581*	0.4556	0.4672	0.4541	
		Kor	1.0000	**	**	0.9962*	0.9957	0.9749	0.9938	
	10%	Kov	4.0539	**	4.1951	4.0239*	4.0235	3.9817	4.0193	
		R ²	0.0087	**	0.0089*	0.0082	0.0098	0.0049	0.011	
		Oran	0.4570	0.4834	0.4834	0.4834	0.4825	0.4877	0.4595*	
	N=100	1%	Kor	1.0000	**	**	0.9992*	0.9992*	0.9959	0.9936
			Kov	2.9426	**	2.9684	2.9384	2.9381*	2.9325	2.9300
			R ²	0.0040	**	0.0041*	0.0041*	0.0041*	0.0043	0.0044
5%		Oran	0.4982	0.4978	0.4978*	0.4978*	0.4975	0.5011	0.4953	
		Kor	1.0000	**	**	0.9973*	0.9970	0.9895	0.9929	
		Kov	2.9426	**	3.0828	2.9271	2.9262	2.9174	2.9322*	
10%		R ²	0.0040	**	0.0046*	0.0046*	0.0047	0.0049	0.0046*	
		Oran	0.4982	0.5099	0.5099	0.5099	0.5096	0.4753	0.5048*	
		Kor	1.0000	**	**	0.9481	0.9483*	0.8954	0.9441	
		Kov	2.9426	**	2.96426*	2.6454	2.6459	2.8293	2.9467	
		R ²	0.0040	**	0.0032	0.0031	0.0032	0.0176	0.0042*	

Çizelge 5.5. Üstel dağılım MAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri (devamı)

Veri Seti	Örnekleme Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=250	1%		Oran	0.5312	0.5311	0.5311*	0.5311*	0.5309	0.5348	0.5306
			Kor	1.0000	**	**	0.9997*	0.9996	0.9973	0.9997*
			Kov	3.6901	**	3.7178	3.6879	3.6877	3.6897*	3.6883
			R ²	0.0024	**	0.0026*	0.0026*	0.0026*	0.0019	0.0026*
	5%		Oran	0.5312	0.5403	0.5403	0.5403	0.5402	0.5325*	0.5386
			Kor	1.0000	**	**	0.9680*	0.9674	0.9485	0.9587
			Kov	3.6901	**	3.6331*	3.4580	3.4561	3.4520	3.4472
			R ²	0.0024	**	0.0026	0.0025*	0.0027	0.003	0.0026
	10%		Oran	0.5312	0.5356	0.5356	0.5356	0.5355	0.5223	0.5352*
			Kor	1.0000	**	**	0.9543*	0.9539	0.8764	0.9508
			Kov	3.6901	**	3.7358*	3.3607	3.3598	3.4042	3.3766
			R ²	0.0024	**	0.0020	0.0018	0.0022*	0.0132	0.0028
N=500	1%		Oran	0.4898	0.4898	0.4898*	0.4898*	0.4897	0.4908	0.4908
			Kor	1.0000	**	**	0.9993	0.9994	0.9982	0.9994
			Kov	3.6892	**	3.7220	3.6847	3.6852	3.6844	6867*
			R ²	0.0068	**	0.0066	0.0065	0.0067*	0.0061	0.0065
	5%		Oran	0.4898	0.4901	0.4901	0.4901	0.4897*	0.4920	0.4927
			Kor	1.0000	**	**	0.9834*	0.9833	0.9716	0.9796
			Kov	3.6892	**	3.7564*	3.5682	3.5686	3.5924	3.5766
			R ²	0.0068	**	0.0069*	0.0064	0.0074	0.0052	0.0063
	10%		Oran	0.4898	0.4896	0.4896*	0.4896*	0.4894	0.4938	0.4909
			Kor	1.0000	**	**	0.9689	0.9691*	0.9502	0.9562
			Kov	3.6892	**	3.8498	3.4640	3.4657	3.5058*	3.4518
			R ²	0.0068	**	0.0065	0.0058	0.0072	0.0056	0.0069*
N=1000	1%		Oran	0.5332	0.5331	0.5331*	0.5331*	0.5328	0.5329	0.5326
			Kor	1.0000	**	**	0.9985*	0.9985*	0.9934	0.9964
			Kov	3.6226	**	3.6487	3.6121	3.6121*	3.6075	3.6090
			R ²	0.0019	**	0.0021	0.0020*	0.0021	0.0021	0.0018*
	5%		Oran	0.5332	0.5374	0.5374	0.5374	0.5370	0.5321*	0.5358
			Kor	1.0000	**	**	0.9713*	0.9715	0.9279	.09617
			Kov	3.6226	**	3.5977*	3.4176	3.4187	3.3641	3.4051
			R ²	0.0019	**	0.0015	0.0014	0.0016	0.0013	0.0020*
	10%		Oran	0.5332	0.5417	0.5417	0.5417	0.5410	0.5444	0.5408*
			Kor	1.0000	**	**	0.9461	0.9464*	0.9065	0.9387
			Kov	3.6226	**	3.6037*	3.2430	3.2444	3.2697	3.2542
			R ²	0.0019	**	0.0015	0.0013	0.0017*	0.0008	0.0014

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.5.'de gerçek değerlerle yöntemlere ait değerlerin oranları karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ancak, sırasıyla ortalama atama ile liste silme yöntemlerinin diğer yöntemlerden daha başarılı olduğu, özellikle eksik veri miktarı arttıkça çoklu atamanın etkinliğinin diğer yöntemlerden daha yüksek olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla liste silme ve ortalama atama yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir.

Kovaryansların karşılaştırmasında da tüm yöntemlerin gerçek değerle farkının anlamlı olacak şekilde fazla olmadığı gözlemlenmiştir. Ayrıca liste silme yönteminin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir

Belirlilik katsayısı (Multiple R²) karşılaştırmasında ise tüm yöntemlerin yakınsamasının gerçek değerle anlamlı farklılaşmadığı, sırasıyla çoklu atama ve regresyon ile atama yöntemlerinin diğer yöntemlerden daha etkin oldukları gözlemlenmiştir.

Çizelge 5.6. Üstel dağılım MNAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=50	1%	Oran							
		Kor							
		Kov							
	5%	R ²	0.4570	0.5114	0.5114*	0.5114*	0.5137	0.5221	0.5122
		Oran	1.0000	**	**	0.8169*	0.8008	0.7458	0.8122
		Kor	4.0539	**	2.8210*	2.7058	2.6534	2.4929	2.6903
		Kov	0.0087	**	0.0421	0.0413	0.0428	0.0454*	0.0416
		R ²	0.4570	0.5631	0.5631*	0.5631*	0.5635	0.5752	0.5631*
		Oran	1.0000	**	**	0.6989	0.6965	0.6415	0.7043*
	10%	Kor	4.0539	**	2.1566*	1.9806	1.9736	1.8680	2.0081
		Kov	0.0087	**	0.0077*	0.0055	0.0055	0.012	0.0062
		R ²	0.4982	0.5117	0.5117	0.5117	0.5118	0.5083*	0.5150
Oran		1.0000	**	**	0.9510	0.9501	0.9729*	0.9248	
Kor		2.9426	**	2.6890	2.6618	2.6592	2.7320*	2.5960	
Kov		0.0040	**	0.0049	0.0049	0.0049	0.0045*	0.0054	
N=100	5%	Oran	0.4982	0.5654	0.5654	0.5654	0.5674	0.5641*	0.5651
		Kor	1.0000	**	**	0.7909*	0.7782	0.7779	0.7880
		Kov	2.9426	**	1.9388*	1.8409	1.8137	1.8572	1.8479
		R ²	0.0040	**	0.0263	0.0240*	0.0289	0.0300	0.0288
		Oran	0.4982	0.6372	0.6372	0.6372	0.6378	0.6670	0.6329*
		Kor	1.0000	**	**	0.6261	0.6198	0.4443	0.6489*
	10%	Kov	2.9426	**	1.2833*	1.1536	1.1441	0.8717	1.2028
		R ²	0.0040	**	0.0233	0.0202	0.0268	0.0128*	0.0264
		Oran	0.5312	0.5657	0.5657	0.5657	0.5657	0.5638*	0.5711
		Kor	1.0000	**	**	0.8359	0.8353	0.8427*	0.7795
		Kov	3.6901	**	2.6102	2.5788	2.5771	2.6121*	2.4211
		R ²	0.0024	**	0.0033	0.0032*	0.0033	0.0038	0.0040
N=250	5%	Oran	0.5312	0.6437	0.6437	0.6437	0.6437	0.6433	0.6430*
		Kor	1.0000	**	**	0.6327	0.6326	0.6437*	0.6330
		Kov	3.6901	**	1.5588*	1.4775	1.4772	1.5339	1.4833
		R ²	0.0024	**	0.0001	0.0001*	0.0001*	0.0001*	0.0001*
		Oran	0.5312	0.7215	0.7215	0.7215	0.7219	0.7058*	0.7216
		Kor	1.0000	**	**	0.5119	0.5093	0.5558*	0.5092
	10%	Kov	3.6901	**	1.0749	0.9670	0.9621	1.1281*	0.9700
		R ²	0.0024	**	0.0013	0.0012	0.0014	0.0012	0.0021*

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.6. Üstel dağılım MNAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri (devamı)

Veri Seti	Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=500	1%	Oran	Oran	0.4898	0.5127	0.5127	0.5127	0.5126	0.5120*	0.5137
			Kor	1.0000	**	**	0.8767	0.8773*	0.8729	0.8653
			Kov	3.6892	**	2.8645*	2.8358	2.8379	2.8413	2.8002
			R ²	0.0068	**	0.0073	0.0072	0.0073	0.0063	0.0071*
	5%	Oran	Oran	0.4898	0.5722	0.5722	0.5722	0.5722	0.5702*	0.5751
			Kor	1.0000	**	**	0.7047	0.7041	0.7195*	0.6810
			Kov	3.6892	**	1.9288*	1.8322	1.8314	1.9163	1.7828
			R ²	0.0068	**	0.0123	0.0116	0.0131	0.0111*	0.0130
	10%	Oran	Oran	0.4898	0.6420	0.6420	0.6420	0.6419*	0.6429	0.6434
			Kor	1.0000	**	**	0.5726*	0.5722	0.5612	0.5529
			Kov	3.6892	**	1.3443*	1.2096	1.2090	1.2582	1.1795
			R ²	0.0068	**	0.0046	0.0043	0.0050*	0.0048	0.0048
N=1000	1%	Oran	Oran	0.5332	0.5614	0.5614	0.5614	0.5614	0.5620	0.5611*
			Kor	1.0000	**	**	0.8662	0.8666*	0.8529	0.8656
			Kov	3.6226	**	2.7460*	2.7185	2.7198	2.6952	2.7203
			R ²	0.0019	**	0.0011*	0.0011*	0.0011*	0.0011*	0.0011*
	5%	Oran	Oran	0.5332	0.6368	0.6368	0.6368	0.6368	0.6423	0.6356*
			Kor	1.0000	**	**	0.6698	0.6699*	0.6218	0.6684
			Kov	3.6226	**	1.7111*	1.6254	1.6258	1.5351	1.6318
			R ²	0.0019	**	0.0007	0.0007	0.0008*	0.0008*	0.0008*
	10%	Oran	Oran	0.5332	0.7195	0.7195	0.7195	0.7195	0.7183*	0.7192
			Kor	1.0000	**	**	0.5410	0.5415*	0.5366	0.5406
			Kov	3.6226	**	1.1785*	1.0605	1.0616	1.1057	1.0717
			R ²	0.0019	**	0.0007	0.0007	0.0008*	0.0008*	0.0008*

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.6.'da gerçek değerlerle yöntemlere ait değerlerin oranları karşılaştırılmış ve eksik veri miktarı arttıkça önemli bir farkın olduğu belirlenmiştir. Ayrıca sırasıyla çoklu atama ile tahmini ortalama eşleştirme yöntemlerinin diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olduğu belirlenmiştir. Ayrıca tahmini ortalama eşleştirme yönteminin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Kovaryansların karşılaştırılmasında da gerçek değerlerin yöntemlerin değerleri ile karşılaştırılmasında farkın anlamlı olacak şekilde fazla olduğu, liste silme yönteminin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Belirlilik katsayısı (Multiple R²) karşılaştırılmasında ise tüm yöntemlerin yakınsamasının gerçek değerle anlamlı farklılaştığı, tahmini ortalama eşleştirme ve çoklu atama yöntemlerinin diğer yöntemlerden daha etkin olduğu gözlemlenmiştir.

5.3. Gamma Dağılımı

Gamma dağılımına göre rastgele üretilen, eksik veri mekanizmalarına göre eksiltilebilir ve eksik veri analiz yöntemleri ile tamamlanan verilerin eksik veri mekanizmasına ve eksiltilebilir yüzdelere göre parametrelerinde oluşan durum ile eksik veri analizi yöntemlerinin parametreler üzerindeki etkinliklerinin yakınsama durumları her eksik veri mekanizmasına özel olacak şekilde sırasıyla Çizelge 5.7. Çizelge 5.8. ve Çizelge 5.9.'da gösterilmiş ve çizelgeler ayrı ayrı yorumlanmıştır.

Çizelge 5.7. Gamma dağılımı MCAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti	Örneklem Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama	
N=50	1%	Ölçek Kor Kov R ²	Herhangi Bir Eksik Veri Bulunmadığından İşlem Yapılmamıştır.								
			5%	Ölçek	1.4924	1.4925*	1.4925*	1.4925*	1.4898	1.4904	1.4903
				Kor	1.0000	**	**	0.9464	0.9472	0.9361	0.9523*
	Kov	2.8536		**	2.6648*	2.5561	2.5675	2.5345	2.5787		
	10%	Ölçek Kor Kov R ²	Ölçek	1.4924	1.4844	1.4844	1.4844	1.4871	1.4814	1.4948*	
			Kor	1.0000	**	**	0.9262	0.9233*	0.8640	0.9140	
			Kov	2.8536	**	2.7266*	2.4483	2.5067	2.3210	2.5710	
	N=100	1%	Ölçek Kor Kov R ²	Ölçek	1.5197	1.5191*	1.5191*	1.5191*	1.5191*	1.5149	1.5180
				Kor	1.0000	**	**	0.9995*	0.9995*	0.9579	0.9945
				Kov	2.6156	**	2.6397	2.6130*	2.6130*	2.5912	2.6070
		5%	Ölçek Kor Kov R ²	Ölçek	1.5197	1.5181	1.5181	1.5181	1.5178	1.5125	1.5187*
				Kor	1.0000	**	**	0.9977*	0.9974	0.9652	0.9974
Kov				2.6156	**	2.7421	2.6036	2.6030	2.5839	2.6191*	
10%		Ölçek Kor Kov R ²	Ölçek	1.5197	1.5184	1.5184	1.5184	1.5177	1.5209	1.5195*	
			Kor	1.0000	**	**	0.9874*	0.9869	0.9624	0.9842	
			Kov	2.6156	**	2.8370	2.5504	2.5495	2.5778*	2.5531	
N=250		1%	Ölçek Kor Kov R ²	Ölçek	1.4972	1.4980*	1.4980*	1.4980*	1.4980*	1.4989	1.4981
				Kor	1.0000	**	**	0.9981*	0.9980	0.9922	0.9976
				Kov	2.3092	**	2.3194	2.3008*	2.3006	2.2920	2.2996
	5%	Ölçek Kor Kov R ²	Ölçek	1.4972	1.4978	1.4978	1.4978	1.4980	1.4977*	1.4985	
			Kor	1.0000	**	**	0.9788*	0.9784	0.9563	0.9673	
			Kov	2.3092	**	2.3247*	2.2126	2.2119	2.2103	2.2005	
	10%	Ölçek Kor Kov R ²	Ölçek	1.4972	1.5001*	1.5001*	1.5001*	1.5006	1.5026	1.5006	
			Kor	1.0000	**	**	0.9551*	0.9521	0.9231	0.9438	
			Kov	2.3092	**	2.3420*	2.1069	2.1013	2.1285	2.0993	
			R ²	0.0055	**	0.0102	0.0091	0.0011	0.0074*	0.0110	

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.7. Gamma dağılımı MCAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri (devamı)

Veri Seti	Örnekleme Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama	
N=500	1%	Ölçek	1.4973	1.4970*	1.4970*	1.4970*	1.4970*	1.4964	1.4966		
			Kor	1.0000	**	**	0.9982*	0.9981	0.9940	0.9976	
				Kov	2.2960	**	2.3111	2.2879*	2.2878	2.2836	2.2876
				R ²	0.0071	**	0.0073	0.0073	0.0073	0.0078	0.0072*
	5%	Ölçek	1.4973	1.4968*	1.4968*	1.4968*	1.4967	1.4959	1.4958		
			Kor	1.0000	**	**	0.9801*	0.9801	0.9650	0.9709	
				Kov	2.2960	**	2.3222	2.2059	2.2061	2.2101*	2.2006
				R ²	0.0071	**	0.073	0.0071*	0.0076	0.0064	0.0085
	10%	Ölçek	1.4973	1.5000	1.5000	1.5000	1.5000	1.5009	1.4995*		
			Kor	1.0000	**	**	0.9463	0.9459	0.9059	0.9530*	
				Kov	2.2960	**	2.0947	2.0562	2.0561	2.0855	2.0974*
				R ²	0.0071	**	0.0076	0.0069*	0.0084	0.0077	0.0116
N=1000	1%	Ölçek	1.5006	1.5002	1.5002	1.5002	1.5002	1.4999	1.5003*		
			Kor	1.0000	**	**	0.9911	0.9914*	0.9833	0.9906	
				Kov	2.3427	**	2.3247*	2.3016	2.3022	2.3008	2.3038
				R ²	0.0052	**	0.0052*	0.0045	0.0046	0.0052*	0.0049
	5%	Ölçek	1.5006	1.5000	1.5000	1.5000	1.5000	1.5011	1.5004*		
			Kor	1.0000	**	**	0.9725	0.9731*	0.9473	0.9586	
				Kov	2.3427	**	2.3326*	2.2159	2.2176	2.2321	2.1970
				R ²	0.0052	**	0.0042	0.0039	0.0044*	0.0031	0.0039
	10%	Ölçek	1.5006	1.4989	1.4989	1.4989	1.4989	1.4986	1.4990*		
			Kor	1.0000	**	**	0.9475	0.9481*	0.8896	0.9357	
				Kov	2.3427	**	2.3375*	2.1035	2.1053	2.1052	2.1012
				R ²	0.0052	**	0.0046	0.0041	0.0052*	0.0054	0.0048

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.7.'de gerçek değerlerle yöntemlere ait değerlerin ölçek parametreleri karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ayrıca, çoklu atama yönteminin diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla ortalama atama ve regresyon ile atama yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir.

Kovaryansların karşılaştırılmasında da tüm yöntemlerin gerçek değerle farkının anlamlı olacak şekilde fazla olmadığı gözlemlenmiştir. Ayrıca liste silme yönteminin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir.

Belirlilik katsayısı (Multiple R²) karşılaştırılmasında ise tahmini ortalama eşleştirme yöntemi dışındaki yöntemlerin gerçek değer ile farkının anlamlı olmadığı gözlemlenmiştir. Ayrıca, sırasıyla çoklu atama ve ortalama atama yöntemlerinin diğer yöntemlerden daha etkin oldukları gözlemlenmiştir.

Çizelge 5.8. Gamma dağılımı MAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti	Örnekleme Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama	
N=50	1%	Ölçek Kor Kov R ²	Herhangi Bir Eksik Veri Bulunmadığından İşlem Yapılmamıştır.								
			5%	Ölçek	1.5046	1.5033	1.5033	1.5033	1.5049*	1.5082	1.5053
				Kor	1.0000	**	**	0.9969*	0.9932	0.9835	0.9910
				Kov	1.4784	**	1.5320*	1.4695	1.4695	1.4996	1.4695
	R ²	0.0787		**	0.0861	0.0792*	0.0930	0.0807	0.0965		
	10%	Ölçek	1.5046	1.4966	1.4966	1.4966	1.4982	1.5016	1.5043*		
		Kor	1.0000	**	**	0.9560	0.9446	0.9680*	0.9114		
		Kov	1.4784	**	1.5050*	1.3514	1.3426	1.4031	1.3430		
		R ²	0.0787	**	0.1051	0.0950*	0.1150	0.0958	0.1263		
	N=100	1%	Ölçek Kor Kov R ²	Ölçek	1.5086	1.5076	1.5076*	1.5076*	1.5073	1.5059	1.5063
				Kor	1.0000	**	**	0.9977	0.9958	0.9811	0.9867*
				Kov	1.9250	**	1.9360	1.9165*	1.9133	1.9001	1.9041
R ²				0.0338	**	0.0377	0.0371	0.0383*	0.0430	0.0416	
5%		Ölçek	1.5086	1.5069	1.5069	1.5069	1.5072	1.5018	1.5079*		
		Kor	1.0000	**	**	0.9954*	0.9941	0.9433	0.9830		
		Kov	1.9250	**	2.0091	1.9076*	1.9075	1.9087	1.9011		
		R ²	0.0338	**	0.0364	0.0342*	0.0388	0.0152	0.0426		
10%		Ölçek	1.5086	1.5034	1.5034	1.5034	1.5041*	1.5035	1.5056		
		Kor	1.0000	**	**	0.9539	0.9637	0.9054	0.9680*		
		Kov	1.9250	**	1.9485*	1.7517	1.7718	1.7670	1.7902		
		R ²	0.0338	**	0.0094	0.0075	0.0118	0.0097	0.0153*		
N=250	1%	Ölçek Kor Kov R ²	Ölçek	1.5052	1.5050	1.5050*	1.5050*	1.5049	1.5045	1.5050*	
			Kor	1.0000	**	**	0.9992	0.9993	0.9977	0.9996*	
			Kov	1.9984	**	2.0115	1.9954	1.9956	1.9931	1.9963*	
			R ²	0.0061	**	0.0060	0.0060	0.0060	0.0061*	0.0061*	
	5%	Ölçek	1.5052	1.5022	1.5022	1.5022	1.5022	1.5021	1.5026*		
		Kor	1.0000	**	**	0.9762*	0.9760	0.9651	0.9702		
		Kov	1.9984	**	2.0012*	1.9047	1.9046	1.9131	1.9092		
		R ²	0.0061	**	0.0066	0.0062*	0.0070	0.0069	0.0115		
	10%	Ölçek	1.5052	1.5035	1.5035	1.5035	1.5035	1.5036*	1.5036*		
		Kor	1.0000	**	**	0.9504	0.9508*	0.8811	0.9381		
		Kov	1.9984	**	2.0067*	1.8052	1.8063	1.7582	1.7960		
		R ²	0.0061	**	0.0047	0.0043	0.0050	0.0056*	0.0049		
N=500	1%	Ölçek Kor Kov R ²	Ölçek	1.5013	1.5009	1.5009	1.5009	1.5010*	1.5005	1.5006	
			Kor	1.0000	**	**	0.9990*	0.9990*	0.9963	0.9986	
			Kov	2.1950	**	2.2132	2.1910*	2.1909	2.1887	2.1909	
			R ²	0.0055	**	0.0056	0.0055*	0.0057	0.0059	0.0055*	
	5%	Ölçek	1.5013	1.5008	1.5008	1.5008	1.5009	1.5011*	1.5000		
		Kor	1.0000	**	**	0.9787*	0.9785	0.9683	0.9707		
		Kov	2.1950	**	2.2135	2.1026*	2.1025	2.1200	2.1003		
		R ²	0.0055	**	0.0059	0.0056*	0.0062	0.0045	0.0071		
	10%	Ölçek	1.5013	1.5004	1.5004	1.5004	1.5006*	1.5003	1.5001		
		Kor	1.0000	**	**	0.9599*	0.9596	0.8962	0.9506		
		Kov	2.1950	**	2.2477	2.0225*	2.0227	1.9923	2.0232		
		R ²	0.0055	**	0.0060	0.0053*	0.0067	0.0084	0.0069		
N=1000	1%	Ölçek Kor Kov R ²	Ölçek	1.5006	1.5003	1.5003	1.5003	1.5003	1.5009	1.5004*	
			Kor	1.0000	**	**	0.9981*	0.9980	0.9916	0.9975	
			Kov	2.3427	**	2.3574	2.3338	2.3337	2.3334	2.3358*	
			R ²	0.0052	**	0.0054	0.0053*	0.0055	0.0055	0.0054	
	5%	Ölçek	1.5006	1.4993	1.4993	1.4993	1.4994	1.5004*	1.4997		
		Kor	1.0000	**	**	0.9742	0.9744*	0.9572	0.9571		
		Kov	2.3427	**	2.3406*	2.2235	2.2242	2.2385	2.2298		
		R ²	0.0052	**	0.0046	0.0043	0.0049*	0.0043	0.0048		
	10%	Ölçek	1.5006	1.4970	1.4970	1.4970	1.4977	1.4954	1.4978*		
		Kor	1.0000	**	**	0.9490	0.9489*	0.8876	0.9441		
		Kov	2.3427	**	2.3448*	2.1101	2.1106	2.0940	2.1241		
		R ²	0.0052	**	0.0055*	0.0049	0.0061	0.0041	0.0062		

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.8.'de gerçek değerlerle yöntemlere ait değerlerin ölçek parametreleri karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ayrıca, çoklu atama yönteminin diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla ortalama atama ve regresyon ile atama yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir.

Kovaryansların karşılaştırılmasında da tüm yöntemlerin gerçek değerle farkının anlamlı olacak şekilde fazla olmadığı gözlemlenmiştir. Ayrıca sırasıyla liste silme ve ortalama atama yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir.

Belirlilik katsayısı (Multiple R^2) karşılaştırılmasında ise tahmini ortalama eşleştirme yöntemi dışındaki yöntemlerin gerçek değer ile farkının anlamlı olmadığı gözlemlenmiştir. Tahmini ortalama eşleştirme yönteminin yüksek oranlı eksik verilerde farkının anlamlı olduğu belirlenmiştir. Ayrıca ortalama atama yönteminin diğer yöntemlerden daha etkin olduğu gözlemlenmiştir.

Çizelge 5.9. Gamma dağılımı MNAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri

Veri Seti Örneklem Sayısı	Eksik Veri Yüzdesi	Ölçek Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama	
N=50	1%	Ölçek								
		Kor								
		Kov								
	5%	R^2								
		Ölçek	1.4894	1.4778	1.4778	1.4778	1.4778	1.4778	1.4800*	1.4767
		Kor	1.0000	**	**	0.9256	0.9260*	0.9511	0.9064	
		Kov	1.5038	**	1.3725*	1.2885	1.2905	1.3328	1.2746	
		R^2	0.0532	**	0.0545	0.0525*	0.0566	0.0590	0.0598	
		Ölçek	1.4894	1.4672	1.4672	1.4672	1.4678	1.4578	1.4700*	
10%	Kor	1.0000	**	**	0.8669	0.8729	0.6941	0.8924*		
	Kov	1.5038	**	1.2880*	1.1303	1.1400	0.9776	1.1775		
	R^2	0.0532	**	0.0356	0.0331	0.0387	0.0353	0.0478*		
N=100	1%	Ölçek	1.4857	1.4819	1.4819	1.4819	1.4816	1.4830*	1.4807	
		Kor	1.0000	**	**	0.9581*	0.9520	0.9790	0.9278	
		Kov	1.7461	**	1.6195	1.6031	1.5933	1.6440*	1.5589	
	5%	R^2	0.0277	**	0.0414	0.0410	0.0419	0.0374*	0.0445	
		Ölçek	1.4857	1.4714	1.4714	1.4714	1.4714	1.4713	1.4720*	
		Kor	1.0000	**	**	0.8766	0.8741	0.8307	0.8837*	
		Kov	1.7461	**	1.4133*	1.3420	1.3386	1.3166	1.3594	
		R^2	0.0277	**	0.0398	0.0277*	0.0391	0.0483	0.0380	
		Ölçek	1.4857	1.4598	1.4598	1.4598	1.4594	1.4584	1.4605*	
	10%	Kor	1.0000	**	**	0.7973	0.7888	0.7438	0.8087*	
		Kov	1.7461	**	1.2349*	1.1102	1.0991	1.0925	1.1331	
		R^2	0.0277	**	0.0468	0.0277*	0.0454	0.0571	0.0506	

Çizelge 5.9.Gamma dağılımı MNAR mekanizmasına ve eksik veri yüzdesine göre parametre değerleri (devamı)

Veri Seti	Örnekleme Sayısı	Eksik Veri Yüzdesi	Parametreler	Gerçek Değer	Eksiltilmiş Değer	Liste Silme	Ortalama Atama	Regresyon Atama	Tahmini Ortalama Eşleştirme	Çoklu Atama
N=250	1%	Ölçek	1.5016	1.4972	1.4972	1.4972	1.4972	1.4972	1.4985*	1.4958
		Kor	1.0000	**	**	0.9587	0.9585	0.9771*	0.9265	
		Kov	1.9808	**	1.8431	1.8209	1.8204	1.8647*	1.7712	
			R ²	0.0195	**	0.0209	0.0209	0.0209	0.0203*	2.0213
	5%	Ölçek	1.5016	1.4845	1.4845	1.4845	1.4847	1.4805	1.4848*	
		Kor	1.0000	**	**	0.8702	0.8713*	0.7863	0.8624	
		Kov	1.9808	**	1.5962*	1.5000	1.5030	1.3963	1.4989	
			R ²	0.0195	**	0.0209	0.0196*	0.0222	0.0190	0.0252
	10%	Ölçek	1.5016	1.4753	1.4753	1.4753	1.4753	1.4780*	1.4754	
		Kor	1.0000	**	**	0.8161	0.8144	0.8263*	0.8067	
		Kov	1.9808	**	1.4666*	1.3193	1.3191	1.3822	1.3153	
			R ²	0.0196	**	0.0286	0.0253*	0.0322	0.0254	0.0295
N=500	1%	Ölçek	1.4973	1.4931	1.4931*	1.4931*	1.4930	1.4926	1.4927	
		Kor	1.0000	**	**	0.9599*	0.9595	0.9436	0.9522	
		Kov	2.2960	**	2.1370*	2.1155	2.1150	2.0948	2.1004	
			R ²	0.0071	**	0.0081	0.0080*	0.0082	0.0092	0.0087
	5%	Ölçek	1.4973	1.4801	1.4801*	1.4801*	1.4801*	1.4799	1.4793	
		Kor	1.0000	**	**	0.8653*	0.8652	0.8430	0.8490	
		Kov	2.2960	**	1.8099*	1.7193	1.7196	1.7074	1.6983	
			R ²	0.0071	**	0.0083	0.0078	0.0088	0.0067*	0.0087
	10%	Ölçek	1.4973	1.4629	1.4629	1.4629	1.4629	1.4629	1.4633*	
		Kor	1.0000	**	**	0.7632	0.7634*	0.7273	0.7602	
		Kov	2.2960	**	1.5133*	1.3374	1.3387	1.3566	1.3529	
			R ²	0.0071	**	0.0113	0.0101*	0.0126	0.0142	0.137
N=1000	1%	Ölçek	1.4991	1.4949	1.4949	1.4949	1.4949	1.4953*	1.4950	
		Kor	1.0000	**	**	0.9605	0.9605	0.9639*	0.9606	
		Kov	2.2796	**	2.1244*	2.1032	2.1032	2.1222	2.1062	
			R ²	0.0048	**	0.0052	0.0051	0.0052	0.0053	0.0050*
	5%	Ölçek	1.4991	1.4807	1.4807	1.4807	1.4807	1.4805	1.4814*	
		Kor	1.0000	**	**	0.8584	0.8581	0.8284	0.8624*	
		Kov	2.2796	**	1.7776*	1.6797	1.6794	1.6653	1.6973	
			R ²	0.0048	**	0.0065	0.0063	0.0068	0.0054*	0.0064
	10%	Ölçek	1.4991	1.4677	1.4677	1.4677	1.4676	1.4692*	1.4675	
		Kor	1.0000	**	**	0.7794*	0.7773	0.7648	0.7717	
		Kov	2.2796	**	1.5457*	1.3848	1.3820	1.4254	1.3866	
			R ²	0.0048	**	0.0120	0.0109*	0.0133	0.0110	0.0148

*: Gerçek Değere En Yakın Değer, **: Hesaplanamayan Değer.

Çizelge 5.9.'da gerçek değerlerle yöntemlere ait değerlerin ölçek parametreleri karşılaştırılmış ve aralarında önemli bir farkın olmadığı belirlenmiştir. Ayrıca, sırasıyla çoklu atama ile tahmini ortalama eşleştirme yöntemlerinin diğer yöntemlerden daha başarılı olduğu gözlemlenmiştir.

Korelasyon katsayılarının karşılaştırılmasında gerçek değerle yöntemlere ait değerler arasındaki farkın anlamlı olmadığı, sırasıyla ortalama atama ve çoklu atama yöntemlerinin diğer yöntemlere göre daha etkin oldukları gözlemlenmiştir. Ayrıca tüm yöntemlerin eksik veri miktarı arttıkça farkın anlamlı hale geldiği gözlemlenmiştir.

Kovaryansların karşılaştırmasında da tüm yöntemlerin gerçek değerle farkının anlamlı olacak şekilde fazla olduğu gözlemlenmiştir. Ayrıca liste silme yönteminin diğer yöntemlere göre daha etkin olduğu gözlemlenmiştir

Belirlilik katsayısı (Multiple R^2) karşılaştırmasında ise tüm yöntemlerin gerçek değerle farklılaşmasının anlamlı olacak düzeyde fazla olduğu ve ortalama atamanın diğer yöntemlerden daha etkin olduğu gözlemlenmiştir

5.4. Multinomial Dağılım

Multinomial dağılıma göre rastgele üretilen, eksik veri mekanizmalarına göre eksiltelen ve eksik veri analiz yöntemleri ile tamamlanan verilerin eksik veri mekanizmasına ve eksiltileme yüzdelerine göre parametrelerinde oluşan durum ile eksik veri analizi yöntemlerinin parametreler üzerindeki etkinliklerinin yakınsama durumları incelenmiş olup multinomial dağılımın özelliğinden dolayı regresyon ile atama ve çoklu atamanın kesin sonuçları verdiği görülmüştür.

6. TARTIŞMA VE SONUÇ

Veri setlerinin doğal yapılarından ya da gözlemlerden kaynaklanan nedenlerden dolayı veri setine ait tüm değişkenler için tüm değerlerin her zaman gözlemlenmesi nerdeyse imkânsızdır. Literatürde veri setinde gözlem değerlerinin gözlenememesi durumu eksik veri olarak adlandırılmaktadır. İstatistiksel analiz yapan programlar ve istatistiksel analizler verilerin eksiksiz olması ilkesine dayandığından eksik verilerin varlığını göz ardı etmek, bu verilerin kullanıldığı analizlerin sonuçlarını etkileyecek ve yanlış sonuçların doğmasına sebep olacaktır.

Araştırmacılar günümüze kadar eksik veri probleminin çözümünde birçok yöntem kullanmışlardır. Bu yöntemlerin genel anlamda geleneksel yöntemler ve modern atama yöntemleri olarak iki alt başlıkta toplandığı görülür. Geleneksel yöntemlerde bulunan silme yöntemleri, eksik verinin bulunduğu değişken ya da birimlerin veri setinden silindiği liste boyunca silme ve eksik gözlemlerin göz ardı edilerek gözlemlenen tüm değerler kullanılarak istatistiksel hesaplamaların yapıldığı ikili silme yönteminden oluşmaktadır. Atama yöntemleri de geleneksel atama ve modern atama yöntemlerini kapsayan ve eksik verileri tahmin ederek veri setini tamamlamayı amaçlayan yöntemlerdir. Bu yöntemler sabit bir değer atanması yöntemi, ortalama atama yöntemi, tahmini ortalama eşleştirme yöntemi, regresyonla atama yöntemi, EM atama yöntemi ve çoklu atama yöntemi gibi yöntemlerdir. Eksik veri sorunu çözülürken, öncelikle veride eksik değerleri ortaya çıkaran bir düzeneğin varlığının araştırmacı tarafından araştırılması gerekir. Bu inceleme sonucunda eksik verinin hangi düzeneğe uygun olduğu incelenerek eksik veri yapısı tanımlanır ve uygun çözüm yöntemi belirlenir. Eksik veri mekanizmaları tamamen rastgele olarak eksik (MCAR), rastgele olarak eksik (MAR) ve rastgele olarak eksik değil (MNAR) şeklinde üçe ayrılmıştır.

Allison (2009) ve Buuren (2012)' e göre eksik veri sorununun çözümünde kullanılan istatistiksel yöntemlerin değerlendirilmesinde geçerli üç genel teorik kriter vardır. Bunlar;

- Yöntemin eksik verilerden kaynaklı yanlılığı minimize etmesi, yani ilgilenilen parametreler için elde edilecek değerlerin mümkün oldukça gerçek değerlere yakın olmasını sağlaması gerekir.

- Yöntemin, hâlihazırdaki elverişli verilerin kullanımını maksimize etmesi yani mevcut verilerin maksimum miktarının kullanılmasını ve verilerin mümkün olduğunca azının atılmasını sağlaması gerekir.
- Yöntemin, belirsizliklere yönelik standart hata, güven aralığı ve olasılık değerlerine yönelik titiz kestirimler sağlaması gerekir.

Çalışmamızda kullandığımız eksik verilerin analizine yönelik geleneksel yöntemler, veriler özellikle MCAR olmadığı durumda, yukarıdaki üç teorik kriter üzerinde iyi performans göstermemektedirler. Liste boyunca silme, eksik verileri ele almak için üç kriterden hiç birini iyi karşılamamaktadır, çünkü tarafsız parametre tahminleri üretememektedir, çok fazla toplanmış bilgiyi boşa harcar ve belirsizliğin kesin tahminlerini vermez. Ortalama atama, mevcut verilerden daha fazla yararlanırken, diğer iki teorik değerlendirme kriterini sağlamamaktadır. Regresyon ile atama, yansız parametre tahminleri verir ve mevcut tüm verileri kullanırken, standart hataların küçülmesini (daha az değer vermeyi) sağlayacağından üç kriterden ikisini karşılayıp birini karşılayamamaktadır. Tahminli ortalama eşleştirme ve çoklu atama, yaklaşık olarak tarafsız parametreler ürettikleri, mevcut tüm verileri kullandıkları ve veri MCAR veya MAR olduğu sürece değişkenlik üzerindeki minimize etme etkisine sahip olmadıkları için eksik veri yöntemleri için üç kriteri de karşılamaktadır.

Çalışmamız sonucunda; Tüm dağılımlar için,

Çalışma kapsamında uygulanan eksik veri analiz yöntemlerinin (liste bazında silme, ortalama atama, regresyonla değer atama, tahmini ortalama eşleştirme ve çoklu atama) parametreleri gerçek değere yakınsamasında oldukça yüksek performans sergileyebileceği görülmüş olmakla birlikte eksik verilerle baş etmenin en iyi yolu, eksikliğin olmaması yönünde çaba sarf edilmesidir. Araştırmacıların eksik veri ile karşılaşması durumunda istatistiksel düzeltmelerden önce, eksikliklerin örüntülerine ve nedenlerine ilişkin bilgileri toplaması ve incelemesi gerekmektedir. Eksik veri miktarının fazla olduğu durumlarda, öncelikle eksik verinin tamamlanmaya çalışılması, bu çabanın yetersiz kaldığı durumlarda ise veri toplama sürecinin tekrarlanması düşünülebilir.

Dağılımların hepsinde MCAR ve MAR tipi eksik verilerde kullanılan tüm yöntemlerin gerçek değerler ile yakınsamada etkili olduğu, çoklu atamanın kullanılan diğer yöntemlerden daha etkili olduğu tespit edilmiştir. Bu yüzden MCAR ve MAR tipi verilerde çoklu atama diğer yöntemlere göre daha iyi sonuçlar vermiştir. Tahmini

ortalama eşleştirme yöntemi çoklu atamadan sonra en iyi sonuçları vermiştir. Ulaşılan bu sonuçlar konu ile ilgili olarak yapılan çeşitli araştırmaların sonuçları ile büyük benzerlik göstermektedir (Graham ve ark.,1994; Bernaards ve Sijtsma, 2000; Musil ve ark., 2002; Streiner, 2002; Bal, 2003; Enders, 2004; Peng ve ark., 2006; Cheema, 2012; Kürşad ve Nartgün, 2015; Ser ve Bati, 2015). Özellikle MCAR ve MAR tipi eksik veri yapılarında eksik veri miktarı arttıkça çoklu atama yönteminin tam veri setlerine en yakın değerleri vermesi Twisk ve ark. (2013), Zhu (2014), Leite ve Beretvas (2010) ve Young ve ark.(2011)'nin çalışmalarıyla tutarlılık göstermektedir.

MNAR yapısındaki eksik verilerde sistematik bir eksiklik söz konusu olduğundan bu eksiklikler ihmal edilerek yapılacak tahminlerin sapmalı çıkması Silva ve Zarate, (2014)'in çalışmasıyla tutarlılık göstermiştir. Aynı zamanda, bu veri tipinde de diğer yöntemlere göre çoklu atamanın daha etkili olduğu ortaya çıkmıştır.

Multinomial dağılım, özelliğinden dolayı çok değişkenli olduğundan ve bu değişkenler arasında belli bir bağlantı olduğundan eksik verili durumların hepsinde çoklu atama ve regresyon ile atamanın kesin sonuçları verdiği anlaşılmıştır.

Ayrıca çalışmamızda elde edilen sonuçların farklı yöntemler uygulandığında elde edilecek sonuçların ve verilerin özelliklerine, verilerin miktarına, eksik veri/gözlem sayısına göre farklılıklar gösterebileceği dolayısı ile böylesi durumlarla karşılaşılan araştırmalarda, sonuçların irdelemesinde, özellikle de anlamlılığa yakın durumlarda katı yorumlardan kaçınılması gerektiği anlaşılmaktadır.



KAYNAKLAR

- Akbař, U. Tavřancıl, E. 2015. Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Deęerlendirme Dergisi*, 6 (1): 38-57.
- Allison, P. D., 2000. Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research*, 28: 301-309.
- Allison, P. D., 2003 . Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 4(1): 545-557.
- Allison, Paul D. 2009. *Missing Data, Handbook of Quantitative Methods in Psychology* (Editör: Roger E. Millsap ve Alberto Maydeu-Olivares), Sage Publications, 72-89.
- Allison, Paul D. 2012. *Handling Missing Data by Maximum Likelihood*. SAS Global Forum.32.
- Alpar, R., 2017. *Uygulamalı Çok Deęişkenli İstatistiksel Yöntemler*. Detay Yayıncılık. Ankara. 840.
- Bal, C., 2003. *Çok Gruplu Veri Setlerinde Eksik Gözlem Sorununun Çözümlemesi ve Sağlık Alanında Bir Uygulama* (doktora tezi), Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.
- Baraldi A.N., Enders. C.,2010. An introduction to modern missing data analyses, *Journal of School Psychology*, 48 : 5–37
- Baygöl, A., 2007, *Kayıp Veri Analizinde Sıklıkla Kullanılan Etkin Yöntemlerin Deęerlendirilmesi*, Yüksek Lisans Tezi, İ. Ü. Sağlık Bilimleri Enstitüsü, İstanbul.
- Bergmeir, C., Benitez, J.M., 2012. On the use of cross-validation for time series predictor evaluation, *Inform Sciences*, 191:192-213
- Bernaards, C.A., Sijtsma, K., 2000. Influence of imputation and em methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research*, 35(3): 321-364.
- Buuren, S. V., Oudshoorn K.G., 2011. Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45 (3): 1-67.
- Buuren, S.V., 2012. *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman & Hall/CRC.416.
- Carpita, M., Manisera, M., 2011. On the imputation of missing data in surveys with likert-type scales. *Journal of Classical*, 28: 93-112
- Cheng, T. S., Victoria-Feser, M. P., 2002, High-breakdown estimation of multivariate mean and covariance with missing observations. *British J. Math. Statist. Psych.*, 55: 317–335.
- Cook, R. J., Zeng, L., Yi, G. Y., 2004. Marginal analysis of incomplete longitudinal binary data: a cautionary note on locf imputation. *Biometrics*, 60: 820-828.
- Cumming, P., 2013. Missing data and multiple imputation. *Clinical Review & Education*, 167 (7): 656-661.
- Demir, E., 2013. Kayıp verilerin varlığında çoktan seçmeli testlerde madde ve test parametrelerinin kestirilmesi: SBS örneęi. *Eğitim Bilimleri Arařtırmaları Dergisi*, 3(2): 48-68.
- Demir, E., Parlak, B., 2012. Türkiye’de eğitim arařtırmalarında kayıp veri sorunu. *Eğitimde ve Psikolojide Ölçme ve Deęerlendirme Dergisi*, 3(1): 20-241.

- Demir,E, 2012. Araştırmalarda kayıp veri sorunu, *Çağdaş Eğitim Dergisi Akademik*, **1**(3): 78-87
- Dempster, A. P., Laird N. M., Rubin D. B., 1977. Maximum likelihood estimation from incomplete data the EM algorithm (with Discussion). *Journal of the Royal Statistical Society* , **39** (1): 1–38.
- Donders, A.R.T., Heijden, G.J.M.G., Stijnen, T., Moons, K.G.M., 2006. A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, **59**: 1087-1091
- Downey, R.G., King, C.V. 1998. Missing data in likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, **125** (2): 175-191
- Duncan, T.E., Duncan, S.C. ve Li, F. 1998. A comparison of model and multiple imputation based approaches to longitudinal analyses with partial missingness. *Structural Equation Modeling: A Multidisciplinary Journal*, **5** (1): 1-21
- Enders, C. K., 2010. *Applied Missing Data Analysis*. The Guilford Press, New York,401.
- Enders, C.K., 2004. The impact of missing data on sample reliability estimates: implications for reliability reporting practices. *Educational and Psychological Measurement*, **64**(3): 419-436.
- Genc, S., Boran, F.E., Akay, D., Xu, Z.S., 2010. Interval multiplicative transitivity for consistency, missing values and priority weights of interval fuzzy preference relations, *Inform Sciences*, **180**: 4877-4891.
- Graham, J.W., 2009. Missing Data Analysis: Making it work in the real world. *Annual Review of Psychology*, **60**: 549-576.
- King, G., J. Honaker, A. Joseph,K. Scheve., 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, **95** (1): 49–69.
- Kürşad,M.Ş., Nartgün. Z , (2015). Eksik veri sorununun çözümünde kullanılan farklı yöntemlerin ölçeklerin geçerlik ve güvenilirliği bağlamında karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, **6** (2): 254-267
- Lee, H., Rancourt, E., Srdal, C. E., 1994. Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics* **10**: 231-238.
- Leite, W. ve Beretvas S.N. 2010. The performance of multiple imputation for likert-type items with missing data. *Journal of Modern Applied Statistical Methods*, **9** (1): 64-74.
- Little, R. J. A., Rubin, D. B., 1987. *Statistical Analysis with Missing Data*, John Wiley & Sons, New York. 371.
- Little, R. J. A., Rubin, D. B., 2002. *Statistical Analysis with Missing Data (2nd ed.)*, John Wiley & Sons, New York. 381.
- Little, R.J.A., 1988. Missing data adjustments in large surveys. *Journal of Business & Economic Statistics*, **6** (3): 287-296.
- Liu, G., Gould, A. L., 2002. Comparison of alternative strategies for analysis of longitudinal trials. *Journal of Biopharmaceutical Statistics*, **12**: 207–226.
- Longford, N.T., 2006. *Missing Data and Small-Area Estimation*, Springer, 376 .
- Mallinckrodt, C. H., Clark, W. S., David, S. R., 2001. Accounting for dropout bias using mixed effects models, *Journal of Biopharmaceutical Statistics*, **11**: 9–21.
- Marsh, H. W., 1998. Pairwise deletion for missing data in structural equation models: nonpositive definite matrices, parameter estimates, goodness of fit and adjusted sample sizes. *Structural Equation Modeling* **5** (1): 22–36.

- McKnight, P. M., 2007. *Missing Data : A Gentle Introduction*, The Guilford Press. . New York.251.
- Meng, Z.Q., Shi, Z.Z., 2012. Extended rough set-based attribute reduction in inconsistent incomplete decision systems. *Inform Sciences*, **204**: 44-69.
- Musil,C.M., Warner, C.B, Yobas, P.K., Jones, S.L., 2002. A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, **24** (7):815-829.
- Nartgün, Z.,2015. Kayıp veri sorununun çözümünde kullanılan farklı yöntemlerin farklı kayıp veri koşulları altında ölçeklerin psikometrik nitelikleri ve ölçme sonuçları bağlamında karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, **7**(4): 252-265.
- Nelwamondo, F.V., Marwala, T.,2008. Techniques for handling missing data: applications to online condition monitoring. *International Journal of Innovative Computing*, **4**:13-41.
- Orchard, T., Woodbury, M.A., 1972. A missing information principle: Theory and applications, *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, **1**: 697-715.
- Osborne, J. W., 2013. *Best Practices in Data Cleaning*. Sage Publication, California. 226.
- Pelckmans, K., De Brabanter, J., Suykens, J.A.K., De Moor, B., 2005. Handling missing values in support vector machine classifiers, *Neural Networks*, **18**: 684-692.
- Peng, C.-Y. J., Harwell, M., Liou, S.-M.,Ehman L. H. ,2006. *Advances in Missing Data Methods and Implications for Educational Research*. (Editör: S. Sawilowsky) , Real Data Analysis. Greenwich, CT: Information Age Publishing Inc.31-78.
- Pigott, T. D., 2001). A review of methods for missing data. *Educational Research and Evaluation*, **7**(1): 353-383.
- Praag V.,Dijkstra T. K., Velzen J. V., 1985. Least squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika* **50** (1): 25–36.
- Rubin, D. B., 1976. Inference and Missing Data, *Biometrika*, **63**: 581–592.
- Rubin, D.B. , 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc., New York.
- Satıcı, E., 2009. *Kayıp Gözlem Olması Durumunda Kitle Ortalaması Tahmini* (doktora tezi), H.Ü. Fen Bilimleri Enstitüsü, Ankara.
- Satıcı, E., Kadılar, C., 2009. Kayıp gözlem olduğunda kitle ortalamasının tahmini. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, **10** (2): 549-556.
- Schafer, J. L., J. W. Graham. 2002. Missing data: Our view of the state of the art. *Psychological Methods* **7** (2): 147–177.
- Schafer, J. L., N. Schenker., 2000. Inference with imputed conditional means. *Journal of the American Statistical Association*, **95** (449): 144–154.
- Schafer, J. L.,1999. Multiple imputation: a primer. *Statistical Methods On Medical Research*, **8**(1): 3-15.
- Schaffer, J.L., 1997. *Analysis of Incomplete Multivariate Data*, Chapman&Hall, London. 444.
- Sehgal, M.S., Gondal, L., Dooley, L.S., 2005. Collateral missing value imputation: A new robust missing value estimation algorithm for microarray data, *Bioinformatics*, **21**: 2417-2423,

- Ser, G., Bati C.T., 2015., Eksik veri analizinde çoklu atama yönteminin değerlendirilmesi: Hayvancılıkta tekrarlı ölçüm verisi üzerine bir uygulama. *Türk Tarım – Gıda Bilim ve Teknoloji Dergisi*, **3** (12): 926-932.
- Silva, L., Zarate, L., 2014. A brief review of the main approaches for treatment of missing data. *Intelligent Data Analysis*, 1177–1198.
- Streiner, D.V., 2002. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Research methods in Psychiatry*, **47**: 68-75.
- Şeker Ş. E., Eşmekaya E., 2017. Eksik verilerin tamamlanması (imputation) *YBS Ansiklopedi*, **4**(3):10-17
- Toka.O., 2012. **Kayıp Veri Durumunda Sağlam Kestirim** (yüksek lisans tezi) Hacettepe Üniversitesi Fen Bilimleri Enstitüsü. Ankara.
- Twisk J, De Boer M, De Vente W, Heymans M. 2013. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *J. Clin. Epidemiol.*, **66**: 1022-1028.
- Van der Ark, L. A., Vermunt, J. K., 2010. New Developments in Missing Data Analysis. *Methodology*, **6**(1), 1-2
- Wilkinson, L., 1999. Statistical methods in psychology journals: guidelines and explanations. *American Psychologist, Task Force on Statistical Inference*, **54**: 594–604.
- Young, W., Weckman, G., Holland, W. 2011. A survey of methodologies for the treatment of missing values within datasets: Limitations and benefits, *Theoretical Issues in Ergonomics Science*, **12** (1): 15-43.
- Zhu X. 2014. Comparison of four methods for handling missing data in longitudinal data analysis through a simulation study. *Open Journal of Statistics*, **4**: 933-944.

ÖZ GEÇMİŞ

Şırnak İdil’de 1979 yılında doğdu. İlkokul ve ortaokulu İdil’de, liseyi Diyarbakır’da tamamladı. 1997 yılında girmiş olduğu İnönü Üniversitesi Matematik Öğretmenliği Bölümünden 2002 yılında mezun oldu. Aynı yılda başladığı öğretmenlik hayatına halen devam etmektedir. 2017 yılında Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda tezli yüksek lisansa başladı. Evli ve üç çocuk babasıdır.

