

**T.C.**  
**ABANT İZZET BAYSAL ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**EĞİTİM BİLİMLERİ ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**SIKLIKLA KULLANILAN KAYIP VERİ YÖNTEMLERİNİN**  
**BETİMSEL İSTATİSTİK, GÜVENİRLİK VE GEÇERLİK**  
**AÇISINDAN KARŞILAŞTIRILMASI**

**Yüksek Lisans Tezi**

**Hazırlayan**  
**Merve Şahin KÜRŞAD**

**Danışman**  
**Doç. Dr. Zekeriya NARTGÜN**

**BOLU, ARALIK, 2014**

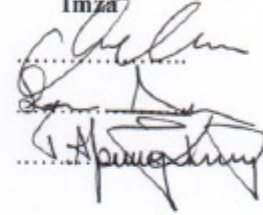
## YÜKSEK LİSANS TEZİ ONAY FORMU

Merve Şahin Kürşad tarafından hazırlanan "Sıklıkla Kullanılan Kayıp Veri Yöntemlerinin Betimsel İstatistik, Güvenirlik ve Geçerlik Açısından Karşılaştırılması" başlıklı çalışma jüri tarafından Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir. 15.11.2014

### Akademik Unvan ve Adı Soyadı

Üye (Tez Danışmanı) : Doç. Dr. Zekeriya NARTGÜN  
Üye : Prof. Dr. Soner DURMUŞ  
Üye : Yrd. Doç. Dr. İ. Alper KÖSE

İmza



**Eğitim Bilimleri Enstitüsünün Onayı**

**Doç Dr. Türkan ARGON**

**Eğitim Bilimleri Enstitüsü Müdürü**

## ETİK İLKELERE UYULDUĐUNA İLİŐKİN BEYAN

Yüksek Lisans Tezi olarak sunduđum, “**Sıklıkla Kullanılan Kayıp Veri Yöntemlerinin Betimsel İstatistik, Güvenirlik ve Geçerlik Açısından Karşılaştırılması**” başlıklı çalışmanın yazılmasında bilimsel ve etik kurallara uyduđumu, başkalarının eserlerinden yararlanılması durumunda atıfta bulunduđumu, kullanılan verilerde herhangi bir tahrifat yapmadıđımı, tezin tamamının ya da bir kısmının bu üniversite veya başka bir üniversitede bir tez çalışması olarak sunulmadıđını beyan ederim. 15/11/2014

Merve Şahin Kürşad

Canım aileme...

## TEŞEKKÜR

Tez yazım sürecinde her türlü sorumu sabırla ve hoşgörü ile cevaplayan, insani, ahlaki ve iş değerleriyle de örnek aldığım değerli hocam ve tez danışmanım Doç. Dr. Zekeriya Nartgün'e,

Her konuda görüşlerine sıklıkla başvurduğum ve kaynak konusunda desteklerini esirgemeyen değerli hocalarım Yrd. Doç. Dr. İ. Alper Köse ve Yrd. Doç. Dr. Sevilay Kilmen'e, her konuda destek ve moral olan değerli hocam Doç. Dr. Şenay Nartgün'e,

Hayatımın her anında maddi, manevi destek olan, her anımda yanımda olan ve beni bugünlere getiren anneme, babama ve kardeşime, bu süreçte sabırla yardımcı olan eşim Servet Meriç Kürşad'a, tez yazım sürecinde desteklerini esirgemeyen arkadaşlarım Arş. Gör. İbrahim Uysal, Arş.Gör. Elif Yazıcı, Arş.Gör. Emine Tunç ve Arş. Gör. Güler Duman'a, tezimi dil ve biçim bakımından inceleyen Caner Canbulat'a, manevi yönden destek olan ve beni yalnız bırakmayan Hatice Begde ve Celal Begde'ye, kısa süre de olsa stres topum olan canım kedim Siyami'ye,

Eğitim hayatım boyunca bugünlere gelmemi ve gelişmemi sağlayan tüm hocalarıma çok teşekkür ederim.

Merve Şahin Kürşad

## İÇİNDEKİLER

<b>ETİK İLKELERE UYULDUĞUNA İLİŞKİN BEYAN</b> .....	iii
<b>İTHAF</b> .....	iv
<b>TEŞEKKÜR</b> .....	v
<b>İÇİNDEKİLER</b> .....	vi
<b>TABLolar DİZİNİ</b> .....	viii
<b>ŞEKİLLER DİZİNİ</b> .....	x
<b>KISALTMALAR DİZİNİ</b> .....	xi
<b>ÖZET</b> .....	xii
<b>ABSTRACT</b> .....	xiv

### BÖLÜM I

1. Giriş .....	1
1.1. Problem Durumu .....	1
1.2. Problem Cümlesi .....	4
1.3. Araştırmanın Amacı .....	5
1.4. Araştırmanın Önemi .....	5
1.5. Sınırlılıklar .....	6
1.6. Tanımlar .....	6

### BÖLÜM II

2. Kuramsal Temeller ve Literatür Taraması .....	7
2.1. Kayıp Veri ile İlgili Genel Bilgiler .....	7
2.2. Kayıp Veri Mekanizmaları .....	8
2.3. Kayıp Veri Yöntemleri .....	13
2.4. Betimsel İstatistikler .....	33
2.5. Güvenirlilik ve Geçerlik .....	34
2.6. İlgili Araştırmalar .....	35

**BÖLÜM III**

3. Yöntem.....	39
3.1. Araştırma Modeli.....	39
3.2. Veri Toplama Aracı ve Veriler.....	39
3.3. Çalışma Grupları ve Verilerin Düzenlenmesi .....	40
3.4. Verilerin Analizi .....	41
3.5. Çalışma Kapsamında Ele Alınan Kayıp Veri Yöntemleri.....	49

**BÖLÜM IV**

4. Bulgular.....	51
4.1. Birinci Alt Probleme İlişkin Bulgular .....	51
4.2. İkinci Alt Probleme İlişkin Bulgular .....	64
4.3. Üçüncü Alt Probleme İlişkin Bulgular .....	70

**BÖLÜM V**

5.Sonuç ve Öneriler.....	86
5.1. Sonuç .....	86
5.2. Öneriler.....	95

<b>KAYNAKÇA .....</b>	<b>96</b>
-----------------------	-----------

<b>ÖZGEÇMİŞ .....</b>	<b>110</b>
-----------------------	------------

## TABLOLAR DİZİNİ

<b>Tablo 3.1.</b> Yüzdelerle göre kayıp veri miktarları.....	41
<b>Tablo 3.2.</b> Farklı oranlardaki kayıp veri miktarına göre toplam puanların dağılımına yönelik betimsel istatistikler.....	42
<b>Tablo 3.3.</b> Little'ın TROK testi sonuçları.....	46
<b>Tablo 3.4.</b> Tam veri setleri için güvenilirlik katsayıları.....	47
<b>Tablo 3.5.</b> Tam veri setleri için açımlayıcı faktör analizi sonuçları.....	48
<b>Tablo 4.1.</b> 200 kişilik veri seti için betimsel istatistik sonuçları.....	52
<b>Tablo 4.2.</b> 200 kişilik veri seti için t-testi sonuçları.....	55
<b>Tablo 4.3.</b> 500 kişilik veri seti için betimsel istatistik sonuçları.....	56
<b>Tablo 4.4.</b> 500 kişilik veri seti için t-testi sonuçları.....	59
<b>Tablo 4.5.</b> 1000 kişilik veri seti için betimsel istatistik sonuçları.....	60
<b>Tablo 4.6.</b> 1000 kişilik veri seti için t-testi sonuçları.....	63
<b>Tablo 4.7.</b> Farklı oranlarda kayıp veri içeren 200 kişilik veri seti için güvenilirlik katsayıları ve Fisher z testi sonuçları.....	65
<b>Tablo 4.8.</b> Farklı oranlarda kayıp veri içeren 500 kişilik veri seti için güvenilirlik katsayıları ve Fisher z testi sonuçları.....	67
<b>Tablo 4.9.</b> Farklı oranlarda kayıp veri içeren 1000 kişilik veri seti için güvenilirlik katsayıları ve Fisher z testi sonuçları.....	68
<b>Tablo 4.10.</b> 200 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	71
<b>Tablo 4.11.</b> 200 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	73
<b>Tablo 4.12.</b> 200 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	74

<b>Tablo 4.13.</b> 500 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	76
<b>Tablo 4.14.</b> 500 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	78
<b>Tablo 4.15.</b> 500 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	79
<b>Tablo 4.16.</b> 1000 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	81
<b>Tablo 4.17.</b> 1000 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	82
<b>Tablo 4.18.</b> 1000 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları.....	84

## ŞEKİLLER DİZİNİ

<b>Şekil 3.1.</b> Farklı oranlarda kayıp veri içeren 200 kişilik veri seti için histogram grafikleri.....	43
<b>Şekil 3.2.</b> Farklı oranlarda kayıp veri içeren 500 kişilik veri seti için histogram grafikleri.....	44
<b>Şekil 3.3.</b> Farklı oranlarda kayıp veri içeren 1000 kişilik veri seti için histogram grafikleri.....	45

**KISALTMALAR DİZİNİ**

<b>TROK</b>	: Tamamıyla Rassal Olarak Kayıp
<b>ROK</b>	: Rassal Olarak Kayıp
<b>IE</b>	: İhmal Edilemez Kayıp
<b>SO</b>	: Seri Ortalaması
<b>YNO</b>	: Yakın Noktaların Ortalaması
<b>YNM</b>	: Yakın Noktaların Medyanı
<b>DDK</b>	: Doğrusal Değer Kestirimi
<b>NDE</b>	: Noktanın Doğrusal Eğimi
<b>LBS</b>	: Liste Bazında Silme
<b>ÇBS</b>	: Çiftler Bazında Silme;
<b>OA</b>	: Ortalama Atama
<b>BM</b>	: Beklenti Maksimizasyonu
<b>RA</b>	: Regresyon Ataması
<b>EÇO</b>	: En Çok Olabilirlik
<b>ÇA</b>	: Çoklu Atama
<b>TBEO</b>	: Tam Bilgi En Çok Olabilirlik
<b>HD</b>	: Hot Deck Ataması
<b>CD</b>	: Cold Deck Ataması

**ÖZET****SIKLIKLA KULLANILAN KAYIP VERİ YÖNTEMLERİNİN BETİMSSEL İSTATİSTİK, GÜVENİRLİK VE GEÇERLİK AÇISINDAN KARŞILAŞTIRILMASI****Şahin Kürşad, Merve****Yüksek Lisans Tezi****Eğitim Bilimleri Anabilim Dalı****Ölçme ve Değerlendirme Bilim Dalı****Tez Danışmanı: Doç. Dr. Zekeriya Nartgün****Aralık - 2014, xv + 110 Sayfa**

Kayıp veri sorunu araştırmacıların sıklıkla karşılaştığı sorunlardan biridir. Bu sorun çalışma sonuçlarını olumsuz etkileyerek hatalı çıkarımlar yapılmasına neden olmaktadır. Bundan dolayı kayıp veri problemine çözüm olarak farklı kayıp veri yöntemleri geliştirilmiştir. Örneklem büyüklüğü, kayıp veri miktarı, kayıp veri mekanizması vb. etkenlere bağlı olarak kayıp veriyi tamamlamada hangi yöntemin kullanılacağı değişkenlik göstermektedir.

Temel araştırma niteliği taşıyan bu çalışmada farklı örneklem büyüklüğü ve farklı oranlarda kayıp veri içeren veri setleri 9 farklı kayıp veri yöntemi ile yeniden yapılandırılmış ve tam veri setleri ile betimsel istatistikler, güvenilirlik ve geçerlik bakımından karşılaştırılarak hangi koşullar altında hangi yöntemlerin kullanılabilir olduğuna ilişkin çıkarımlarda bulunmak amaçlanmıştır. Araştırmada Uluslararası Öğrenci Değerlendirme Programı (PISA) verilerinden yararlanılmıştır. Veriler için PISA 2012 Türkiye örnekleme ve öğrenci ölçeklerinden normal dağılıma sahip, tek faktörlü “Matematik Çalışma Etiği (Math Work Ethic)” ölçeği seçilerek, analizler için rastgele 200, 500 ve 1000 kişilik çalışma grupları oluşturulmuştur. Tam veri setleri üzerinde, tamamıyla rassal olarak kayıp mekanizması altında %5, %10 ve %20

oranlarında silme işlemi gerçekleştirilmiştir. Kayıp verilerin yeniden yapılandırılması aşamasında seri ortalaması, yakın noktaların ortalaması, yakın noktaların medyanı, doğrusal değer kestirimi, noktanın doğrusal eğimi, liste bazında silme, beklenti maksimizasyonu, regresyon ataması ve çoklu atama yöntemleri kullanılmıştır.

Kayıp veri yöntemleriyle karşılaştırmada tam veri setlerinden elde edilen betimsel istatistik, güvenilirlik ve geçerliğe ilişkin değerler referans değerler olarak kullanılmıştır. Kayıp veri yöntemleri kullanılarak yapılandırılan yeni verilerden elde edilen değerler ile referans değerler karşılaştırılarak hangi yöntemin hangi koşullar altında diğerlerine göre daha kullanılabilir olduğuna ilişkin çıkarımlarda bulunulmuştur. Karşılaştırmalar bazı durumlar için betimsel düzeyde, bazı durumlar için t testi ve Fisher'in z testi kullanılmak suretiyle gerçekleştirilmiştir.

Araştırma sonuçlarına göre farklı büyüklük ve farklı kayıp veri oranlarında, liste bazında silme yöntemi için elde edilen değerler, genel olarak tam veri setlerinden elde edilen değerlere en az benzerlik gösteren değerlerdir. Yaklaşık değer atama yöntemleri için elde edilen değerler kayıp veri oranının düşük olduğu durumlarda genel olarak tam veri setlerinden elde edilen değerlere yakın veya aynı değerleri verirken, tam veri setlerinden elde edilen değerlere çok yakın değer veren yöntemler beklenti maksimizasyonu, regresyon ataması ve çoklu atama yöntemleri olmuştur. Belirtilen yöntemler içerisinde özellikle çoklu atama yöntemi için elde edilen değerlerin, diğer yöntemlere göre, tam veri setlerinden elde edilen değerlere daha yakın sonuçlar verdiği görülmüştür. Betimsel karşılaştırmalarda bazı yöntemlerin diğerlerine göre kullanılabilirlik bakımından ön plana çıktığı görülse de yöntemler arasında büyük benzerliklerin olduğu sonucuna ulaşılmıştır.

**Anahtar Kelimeler:** Kayıp veri yöntemleri, betimsel istatistik, geçerlik, güvenilirlik

**ABSTRACT****EVALUATION OF COMMONLY USED MISSING DATA METHODS IN  
TERMS OF DESCRIPTIVE STATISTICS, RELIABILITY AND VALIDITY****Şahin Kürşad, Merve****M. Sc. Thesis****Department of Educational Sciences****Measurement and Evaluation in Education****Supervisor: Assoc. Prof. Dr. Zekeriya Nartgün****December - 2014, xv + 110 Pages**

Missing data is often encountered by researchers. This problem negatively effects the results of researches and causes erroneous inferences. As a solution to this problem different missing data methods were developed. These methods which are used to complete the missing data differ depending on size of sample, quantity of missing data, mechanism of missing data etc.

This research is a basic research and in which data sets with different sample size and different amount of missing data were used. The purpose of this study is to define the suitable methods for different conditions by comparing the complete data sets with data sets which are applied 9 different missing data methods, in terms of descriptive statistics, reliability and validity. For this research, Programme for International Students' Assesment (PISA) data were used. PISA 2012 Turkey sample and "Math Work Ethic" scale which was normally distributed and one factored, was selected and data sets which contained 200, 500 and 1000 data were formed at random. Then by using completely missing at random mechanism, %5, %10 and %20 of data were deleted from each data set. In order to complete these missing data, series mean, mean of nearby points, median of nearby points, linear interpolation, linear trend at

point, listwise deletion, expectation maksimization, regression imputation and multiple imputation techniques were used.

When comparing missing data methods, values obtained from descriptive statistics, reliability and validity were used as referance values. New data values which has been structured with missing data methods compared to referance values in order to make an inference about which method is suitable for different conditions. Some conditiones were compared in descriptive level, some conditiones were compared in terms of t-test and Fisher z test.

The results of the study revealed that for different size and missing data rate, listwise deletion method values had the least similarity with the values obtained from complete data sets. Approximate value imputation methods (estimation) values had results which was close to the complete data set values or same with it when the quantity of missing data was low. The methods which had the closest values to values obtained from the complete data set were expectation maximization, regression imputation and multiple imputation. Multiple imputation method outperformed compared to the others. Although, in descriptive comparisons it was found that some methods were more at the forefront, there was no statistically significant difference between methods according to t-test and Fisher z test.

**Keywords:** Missing data methods, descriptive statistics, validity, reliability.

# BÖLÜM I

## 1. Giriş

Bu bölümde araştırmanın problem durumu, problem cümlesi, alt problemleri, amaç, önem, sınırlılıklar ve tanımlar ele alınmıştır.

### 1.1. Problem Durumu

Yapılan birçok araştırmada, toplanan verilerde eksiklerle karşılaşmaktadır. Verilerde bulunan bu eksikler kayıp veri veya kayıp değer olarak adlandırılır. Bu durum istatistiksel analizlerde çoğu araştırmacının karşılaştığı önemli bir sorundur (Çokluk ve Kayrı, 2011). Örneğin, bilindiği üzere, psikolojik değerlendirmenin temeli gözlenen verilerden hareketle örtük değişkenler hakkında çıkarımlarda bulunmaktır. Ancak gözlenen verilerde bulunacak eksikler, örtük değişkenler hakkında çıkarımda bulunulmasını da zorlaştıracaktır (Hohensinn ve Kubinger, 2011). Bundan dolayı araştırmacılardan kayıp veri sorununu ya baştan engellemeleri ya da bu sorunla karşılaştıklarında olabildiğince düzeltmeleri beklenir (McKnight, McKnight, Sidani ve Figueredo, 2007).

Kayıp veri sorunuyla en çok sosyal bilimler (Vansteelandt, Carpenter, Kenward, 2010) ve kişisel davranışların ölçüldüğü alanlarda karşılaşmaktadır (Ginkel, Sijtsma, Van der Ark, Vermunt, 2010). Uzun anketlerde soruların atlanması, deneysel çalışmalarda verilerin kaydedilmemesi gibi mekanik hatalar, seksüel davranışlar vb. hassas konularda araştırma yapma (Field, 2009), soruların dikkatsizlik sonucu veya cevabı bilinmediği için boş bırakılması (Finch ve Margraf, 2008) gibi sebepler kayıp veri sorununun başlıca sebepleri arasındadır (Sijtsma ve van der Ark, 2003). Brown ve

Kros (2003) bu sebepleri 3 grupta sınıflandırmıştır. Bunlardan birisi süreçle ilgili nedenlerdir. Süreçle ilgili nedenler veri girişi sırasında karşılaşılan sorunları kapsamaktadır. Bu sorunlar verilerin girilmemesi ve veri gruplaması sırasında karşılaşılan problem durumlarını içerir. Bir diğer neden ise cevaplamayı reddetme ile ilgili nedenlerdir. Böyle durumlar, bazı soruların ağır veya hassas konularla ilgili olmasından kaynaklanmaktadır. Üçüncü neden ise herhangi bir durum, grup veya konuyla ilgisi olmayan soruların sorulmasıdır, yani soruların özel bir gruba sorulması yerine daha geniş bir kitleye sorulmasından kaynaklanan kayıp veri durumudur. Örneğin evli olmayan bir gruba kaç yıllık evlisiniz gibi bir soru sorulması kayıp veriye neden olacaktır.

Araştırmalarda yer alan kayıp veriler, bilgi eksikliğini temsil eder, dolayısıyla da bilgi kaybına neden olurlar (Bal, 2003). Araştırmalarda, katılımcılar tüm soruları cevaplandırmaları için bilgilendirilse dahi çalışmalarda kayıp veri durumu ile karşılaşılmaktadır (Ginkel, Sjitsma, Van der Ark ve Vermunt, 2007). İstatistiksel analizler için gerekli olan paket programlar tam veri setlerine göre düzenlendiği için, toplanan verilerde bulunan kayıplar istatistiksel analizlerde önemli sorunlara neden olmaktadır (Bal, 2003). Kayıp verilerden kaynaklanan problemlerin yaşanmaması için araştırmacıların tam veri setleri ile çalışması gerekmektedir.

McKnight ve diğerleri (2007)'ye göre kayıp veriler çalışma sonuçlarını şu şekilde etkilemektedir. Eğer kayıp veri miktarı fazlaysa elde edilen sonuçların güvenilirliği, genellenebilirliği ve yapılacak istatistiksel çıkarımlar bundan önemli derecede etkilenecektir. Bunun sonucunda da yapılan istatistiksel çıkarımlar yanıltıcı olacaktır. Ayrıca kayıp veriler çalışmanın geçerliğini de olumsuz yönde etkileyecektir.

İlgili alanyazında yukarıda belirtilen hususların yanı sıra kayıp verilerin ölçme sonuçlarının ortalaması, standart sapması, basıklık ve çarpıklık değerleri gibi istatistikleri de etkilediğini göstermektedir (Bal, 2003; Demir, 2013). Alanyazında kayıp verilere herhangi bir işlem uygulamadan yapılan betimsel istatistik, güvenilirlik ve geçerlik kestirimlerinin kayıp verinin niteliğine ve miktarına bağlı olarak yanlılık ve kestirim hataları üretebileceği belirtilmektedir (Demir ve Parlak, 2012).

Kayıp veri ile karşılaşılan durumlarda araştırmacılar genellikle kayıp veri içeren durumları analiz dışı bırakmayı tercih etmektedirler. Ancak bunun yapılabilmesi için öncelikle kayıp veriye neyin neden olduğunun anlaşılması gerekmektedir (Demir ve Parlak, 2012; Yuan ve Lu, 2008). Çünkü kayıp veriler farklı nedenlerle ve farklı örüntülerle ortaya çıkmaktadır. Ayrıca, kayıp veri mekanizmasının belirlenmesi, kayıp veri sorunuyla baş etmede hangi kayıp veri yönteminin kullanılacağını belirlemek açısından da önemlidir (Allison, 2003).

Araştırmalarda karşılaşılan kayıp veri problemi için çeşitli çözüm ve atama yöntemleri geliştirilmiştir. Kayıp verilerle karşılaşıldığı durumlarda kayıp veri ile analize devam etme, eksik gözlemleri analiz dışı bırakma, eksik gözlemler yerine veri atama veya çeşitli istatistiksel yöntemlerle eksik verileri tamamlama gibi yöntemler geliştirilmiştir (Bal, 2003; Carpita ve Manisera, 2011; Duncan, Duncan ve Li, 1998; Downey ve King, 1998; Little, 1988). Bu yöntemler içerisinde araştırmalarda en çok tercih edilenler liste bazında silme ve çiftler bazında silme gibi eksik verileri analiz dışı bırakma yöntemleridir. Ancak yapılan çalışmalar bu yöntemlerin örnekleme kayba, güvenilirlikte azalmaya, tahminlerde yanlılığa neden olduğunu (Allison, 2009; Cumming, 2013; Satıcı ve Kadılar, 2009; Oğuzlar, 2001; Van Der Ark ve Vermunt, 2010) ve yanlılıktan kaynaklı olarak da örneklemin evreni temsil etme derecesinin düştüğünü göstermektedir (Demir ve Parlak, 2012, Little, 1988). Belirtilen bu sebeplerden dolayı, son yıllarda, bu yöntemlerin yerine, beklenti maksimizasyonu ve çoklu atama gibi “modern” yöntemler önerilmektedir. Çünkü bu yöntemler, silme yöntemleri gibi geleneksel kayıp veri yöntemlerinin aksine, yanlılığın azaltılması, etkili parametre tahminlerinin yapılması ve daha büyük istatistiksel gücün sağlanması hususunda daha etkili sonuçlar vermektedir (Enders, 2013).

Kayıp veriler için birçok farklı yöntem ve her yöntemin kayıp veriyi tahmin etme konusunda kendine göre farklı etkileri vardır. Kayıp veriler çalışma sonuçlarını farklı şekillerde etkilediği için, araştırmacıların örneklem büyüklüğü, kayıp veri oranı vb. faktörleri dikkate alarak yöntem seçimi yapması gerekmektedir (Cheema, 2012). Kayıp verilerle çalışmaya devam edildiği veya uygun yöntem seçilmediği takdirde

çalışma sonuçları bu durumdan olumsuz yönde etkilenecektir (Ginkel ve diğerleri, 2007).

Bütün bu bilgilerden hareketle bu araştırmada, farklı kayıp veri yöntemlerinin betimsel istatistikler, güvenilirlik, geçerlik üzerindeki etkileri, normal dağılım ve tek faktörlü yapı bağlamında, farklı örneklem büyüklükleri ve farklı oranlardaki kayıp veri yüzdesi altında incelenerek karşılaştırılmıştır.

## 1.2. Problem Cümlesi

Normal dağılıma sahip, tek faktörlü, farklı büyüklükteki tam veri setleri ile değişen oranlarda kayıp veri içeren veri setlerine uygulanan çeşitli kayıp veri yöntemleri ile elde edilen yeni veri setleri arasında betimsel istatistik, güvenilirlik ve geçerlik bakımından manidar farklılık var mıdır?

### 1.2.1. Alt problemler

1. a. Normal dağılıma sahip tek faktörlü tam veri setleri ( $n=200$ ,  $n=500$ ,  $n=1000$ ) ile belirli oranlarda kayıp veri içeren (%5, %10, %20) veri setlerine uygulanan kayıp veri yöntemleri ile elde edilen yeni veri setleri arasında betimsel istatistik değerleri (ortalama, standart sapma, çarpıklık, basıklık) bakımından ne düzeyde farklılıklar görülmektedir?

1. b. Normal dağılıma sahip tek faktörlü tam veri setleri ( $n=200$ ,  $n=500$ ,  $n=1000$ ) ile belirli oranlarda kayıp veri içeren (%5, %10, %20) veri setlerine uygulanan kayıp veri yöntemleri ile elde edilen yeni veri setleri arasında hesaplanan ortalama değerleri bakımından manidar bir farklılık var mıdır?

2. Normal dağılıma sahip tek faktörlü tam veri setleri ( $n=200$ ,  $n=500$ ,  $n=1000$ ) ile belirli oranlarda kayıp veri içeren (%5, %10, %20) veri setlerine

uygulanan kayıp veri yöntemleri ile elde edilen yeni veri setleri arasında cronbach alpha değerleri açısından manidar bir farklılık var mıdır?

3. Normal dağılıma sahip tek faktörlü tam veri setleri (n=200, n=500, n=1000) ile belirli oranlarda kayıp veri içeren (%5, %10, %20) veri setlerine uygulanan kayıp veri yöntemleri ile elde edilen yeni veri setleri açımlayıcı faktör analizi değerleri (madde-faktör yük değerleri, açıklanan varyans, özdeğer) açısından farklılık göstermekte midir?

### 1.3. Araştırmanın Amacı

Bu çalışmanın amacı normal dağılıma sahip tek faktörlü tam veri setleri ile değişen oranlarda kayıp veri içeren farklı büyüklükteki veri setlerine uygulanan çeşitli kayıp veri yöntemleri sonucu elde edilen yeni veri setlerini betimsel istatistik, güvenilirlik ve geçerlik açısından karşılaştırarak, hangi koşullarda hangi yöntemin gerçeğe daha yakın sonuçlar verdiğiine ilişkin çıkarımlarda bulunmaktır. Bu amaç doğrultusunda farklı büyüklükte veri setleri kullanılarak (n=200, n=500, n=1000) belirli oranlarda kayıp değerlerin olduğu koşullara (%5, %10, %20) uygulanan kayıp veri yöntemleri sonucu elde edilen yeni veri setleri, tam veri setleri ile karşılaştırılmıştır. Tam veri setleri tek faktörlü bir yapıyı ölçen bir ölçeğin uygulanması ile elde edilmiş verilerdir. Farklı büyüklük ve oranlarda kayıp veri içeren veri setleri belirtilen bu tam veri setinden türetilmiştir.

### 1.4. Araştırmanın Önemi

Kayıp veri sorunu araştırmalarda karşılaşılan en önemli sorunlardan biridir. Bu çerçevede, kayıp veri sorunu ile başetmek amacıyla farklı kayıp veri yöntemleri geliştirilmiştir. Ancak, geliştirilen yöntemlerin etkisi örneklem büyüklüğü, kayıp veri oranı, kayıp verinin dağılımı vb. etkenlere bağlı olarak değişkenlik göstermektedir. Dolayısıyla uygun kayıp veri yöntemini seçerek kayıp veri sorununu ortadan kaldırmak

güvenirlik, geçerlik, arařtırmada ulařılan sonuçlarının doęruluęu vb. özellikler açısından büyük önem arz etmektedir.

Yukarıda yapılan açıklamalardan hareketle, bu arařtırmada, belirtilen kořullar altında (farklı kayıp veri yöntemleri, farklı örneklem büyüklükleri, farklı oranlarda kayıp veri, tek boyutluluk ve normal dağılım) betimsel istatistikler, güvenilirlik ve geçerlik için ulařılan sonuçların, hem ilgili kiřilere hem de ilgili alanyazımına önemli katkılar saęlayacaęı düşünölmektedir. Kayıp veriler geçerlik ve güvenilirlięi, dolayısıyla da ulařılan sonuçları ve çalıřmanın genellenebilirlięini etkiledięi için bu çalıřmanın özellikle geçerlik ve güvenilirlik ile ilgili çalıřma yapan arařtırmacılara yol gösterici nitelikte olacaęı düşünölmektedir.

#### 1.5. Sınırlılıklar

1. Arařtırmada kullanılan veriler  $n=200$ ,  $n=500$ ,  $n=1000$  birimlik veri setleri ile sınırlıdır.
2. Arařtırma %5, %10, %20 kayıp veri oranları ile sınırlıdır.
3. Arařtırmadaki veri setleri kayıp veri mekanizmalarından TROK yapısı ile sınırlıdır.
4. Arařtırmada kullanılan veri setleri normal dağılım ve tek faktör ile sınırlıdır.
5. Arařtırma yaklařık deęer atama yöntemleri, liste bazında silme, beklenti maksimizasyonu, regresyon ataması ve çoklu atama yöntemleri ile sınırlıdır.

#### 1.6.Tanımlar

Kayıp veri: Veri toplama sürecinde veri toplama aracındaki bazı maddelere katılımcıların çeřitli sebeplerle yanıt vermemesinden kaynaklanan kayıplardır.

Kayıp veri yöntemi: Veri toplama süreci sonunda verilerde kayıpların olması durumunda bu sorunu ortadan kaldırmak amacıyla kullanılan silme ve tamamlama gerektiren yöntemlerdir.

## BÖLÜM II

### 2. Kuramsal Temeller ve Literatür Taraması

Bu bölümde öncelikle kayıp veri yöntemleri ile ilgili genel bilgilere yer verilmiştir. Sonrasında kayıp veri mekanizmaları, kayıp veri yöntemleri, yöntemleri karşılaştırmak için kullanılacak olan betimsel istatistik, güvenilirlik ve geçerlik kavramları ve kayıp verilerle ilgili yapılan çalışmalara yer verilmiştir.

#### 2.1. Kayıp Veri ile İlgili Genel Bilgiler

İstatistiksel analizlerde kayıp veri sorunu 1930'lu yıllara kadar uzanmaktadır. Ancak istatistik programlarındaki yetersizlikten dolayı bu yıllarda, bu konuda çok ilerleme kaydedilememiştir. Bu konuyla ilgili asıl gelişmeler 1980 sonrasında yaşanmıştır (Cheema, 2012). 1980'li yıllarda çalışmalar özellikle beklenti maksimizasyonu algoritması üzerinde yoğunlaşmıştır (Bal, 2003).

Bu sorunla ilgili ilk önemli adımların Little ve Rubin (1987) ve Rubin (1987) tarafından atıldığı görülmektedir. Kayıp verilerle ilgili gerçekleşen bu gelişmelere paralel olarak 1990'larda kayıp veri sorunu istatistiksel programlara yansımaya başlamıştır (Akt. Karal, 2014). 1990'lı yıllarda kayıp veri ile ilgili çalışmaların çoklu atama ve Bayesçi teknikler üzerine yoğunlaşıldığı görülmektedir. Kayıp veri ile ilgili gelişmelere bakıldığında 1980'li yıllardan itibaren kayıp verinin bir problem olmaktan çıkıp kayıp verilerin analiz dışı bırakılması yerine bu verilerin analize alınması için uygun istatistiksel yöntemlerin geliştirildiği görülmektedir (Akt. Bal, 2003). Kayıp veri yöntemleri ile ilgili çalışmalar günümüzde de devam etmekle birlikte bu yöntemlerin çalışma sonuçları üzerinde ne derece etkili olacağı kayıp veri miktarına, kayıp veri

mekanizmasına ve neden kayıp durumu olduğuna bağlı olarak değişmektedir. Ancak bu faktörler içerisinde çalışma sonuçlarına etkisi bakımından kayıp veri mekanizması daha çok ön plana çıkmaktadır (Tabachnick ve Fidell, 2012).

## 2.2. Kayıp Veri Mekanizmaları

Yanıtlayıcıdan (soruları boş bırakma gibi) veya yanıtlayıcı dışında (veriyi yanlış girme gibi) gerçekleşen ve kayıp veriye neden olan sürece kayıp veri süreci denir. Bazı durumlarda kayıp veri ile karşılaşılabilineceği düşünülerek, araştırma kapsamında gerekli önlemler alınabilir (Kaspar, 2011). Ancak, yanıtlayıcıdan kaynaklı durumlarda kayıp veri süreci önceden kestirilemeyebilir. Bu durumda araştırmacının, gerekli analizleri yapabilmesi için kayıp veriye neden olan bir yapının olup olmadığını incelemesi gerekir (Imtiaz, 2007). Bu incelemede araştırmacı, kayıp verinin ne sıklıkta ortaya çıktığı, kayıp verilerin gözlemlere rastgele dağılıp dağılmadığı ve kayıp veriyi oluşturan belli bir yapının olup olmadığını sorularına yanıt arar (Kaspar, 2011).

Kayıp veri mekanizmaları, kayıp verinin dağılımına göre farklı şekillerde sınıflanmaktadır. (Acock, 2005). Bu sınıflandırma, verilerin rastgele mi yoksa birbirleri ile bağlantılı olarak mı kayıp olduğuna göre değişmektedir (Osborne, 2013). Kayıp veri mekanizmaları ile ilgili sınıflama Little ve Rubin (1987) tarafından yapılmıştır. Buna göre kayıp veri mekanizmaları 3 grupta sınıflandırılmıştır; tamamıyla tesadüfi kayıp (Missing completely at random, MCAR, TROK), rassal olarak kayıp (Missing at random, MAR, ROK), ihmal edilemez kayıp (Noignorable, NI, IE).

### 1. Tamamıyla rassal olarak kayıp (Missing completely at random, MCAR, TROK)

Bu yapıda olan kayıp verilerde, verinin kayıp olma olasılığı, veri setinde bulunan diğer değişkenlere ve değişkenin kendisine bağlı değildir (Acock, 2005; Baraldi ve Enders, 2010; Buhi, Goodson ve Neilands, 2008; Cheema, 2012; Dural, 2010). Yani, verilerin kayıp olma durumu şans faktörüne bağlıdır (Sinharay, Stern ve Russel, 2001). Örneğin, bir veri seti X (tutum) ve Y (başarı) olmak üzere iki değişken içeriyor olsun. Y

değişkeninde kayıp veri olması ne X değişkenine ne de Y değişkenine bağlı değil ise Y değişkenindeki kayıp veriler tamamıyla tesadüfi kayıptır. Yani, başarı değişkenindeki kayıp veriler ne tutum değişkeniyle ne de başarı değişkeninin kendisiyle alakalı değildir (Allison,2003). X daima gözlenen, Y ise kayıp veri içeren değişken olmak üzere bu yapının formüsel olarak gösterimi şu şekildedir;

$$P(Y \text{ kayıp} / X, Y) = P(Y \text{ kayıp})$$

Formüle göre, Y'de kayıp olma olasılığı ne gözlenen X değişkenine ne de Y değişkeninin kendisine bağlıdır (Allison, 2009).

## 2. Rassal olarak kayıp (Missing at random, MAR, ROK)

Rassal olarak kayıp, tamamıyla tesadüfi kayıp mekanizmasına göre daha zayıf bir varsayımdır. Bu durumda Y değişkeninde kayıp veri olma olasılığı, X değişkenine bağlıyken, Y değişkeninin kendisine bağlı değildir (Allison, 2009; Allison, 2002, Baraldi ve Enders, 2010, Cheema, 2012; Enders ve Bandalos, 2001; Finch ve Margraf, 2008). Bu kayıp veri mekanizmasında “Rassal” ifadesi geçse de aslında çalışmadaki diğer değişkenlere bağlı olan sistematik bir kayıp veri mekanizması vardır (Baraldi ve Enders, 2010; Dural, 2010). Örneğin, cinsiyet (V1) ve kilo (V2) gibi iki değişken olsun. Kadınlar, erkeklere göre kilolarını saklama eğiliminde oldukları için V2 (kilo) değişkenindeki kayıp veriler V1 (cinsiyet) değişkenine bağlı olarak değişebilecektir. Bu durumda V2 değişkeninde bulunan kayıp veriler rassal olarak kayıptır (Buhi ve diğerleri, 2008). Benzer şekilde X ve Y gibi iki tane test olsun. X sınavından düşük alan bireyler Y sınavına girmeyi reddetmektedirler. Bu durumda Y sınavını tamamlama durumu X sınavından alınan sonuçlara bağlıdır ve Y sınavının kendisi ile ilişkisi yoktur. Bu durumda Y sınav sonuçlarındaki kayıp veriler rassal olarak kayıptır (Enders, 2001 (b)). Yani bu varsayımda kayıp veri olma olasılığı kayıp veya gözlenemeyen verilerden bağımsız olup, diğer gözlenen değişkenlere bağlıdır (Buhi ve diğerleri, 2008; Sinharay ve diğerleri, 2001). ROK mekanizmasında olan kayıp veriler ihmal edilebilir ve bu durumda kayıp veri mekanizmasını modellemeye gerek olmadan geçerli parametre tahminleri yapılabilir (Allison, 2002; Allison, 2009). Kayıp verilerin ihmal edilebilirliğinden kasıt kayıp verilerle, seçkisiz olarak belirlenen gözlenen verilerin

örtüşmesidir. Yani, tam veri setleri ile kayıp veri içeren veri setlerinin dağılımları ve bu veriler üzerinden yapılacak çıkarımlar arasında anlamlı bir fark olmaması gerekir (Rubin, 1976). X daima gözlenen, Y ise kayıp veri içeren değişken olmak üzere ROK yapısının formüsel olarak gösterimi şu şekildedir;

$$P(Y \text{ kayıp} / X, Y) = P(Y \text{ kayıp} / X)$$

Formüle göre X ve Y'nin verilmesi durumunda Y'deki kayıp veri olasılığı ile sadece X'in verilmesi durumunda Y'deki kayıp veri olasılığı aynıdır veya başka bir ifade ile Y'deki kayıp veri olasılığı sadece X değişkenine bağlıyken, Y değişkeninin kendisine bağlı değildir (Allison, 2009; Demir ve Parlak, 2012). Bu durum, bir değişkendeki kayıp veri olasılığının, diğer değişkenler kontrol altına alındığında, bu değişkendeki kayıp verilerin kendi değeriyle ilişkisiz olduğunu gösterir (Acock, 2005; Demir ve Parlak, 2012, Enders, 2001 (a)).

### 3. İhmal edilemez kayıp (Nonignorable, NI, IE)

Kayıp verinin en çok problem yarattığı durumdur. İhmal edilemez kayıp mekanizmasında verilerin kayıp olma olasılığı gözlenemeyen yani kayıp verinin yer aldığı değişkenle ilgilidir. Bu durumda olan kayıp veriler ihmal edilemez ve kayıp veri mekanizmasının, iyi bir parametre tahmini yapabilmek için modellenmesi gerekir (Allison, 2002; Allison, 2009; Little ve Rubin, 1987; Streiner, 2002). İhmal edilemez kayıp veri mekanizması, genellikle hangi modelin uygun olacağı konusunda yeterli bilgi sağlamadığı için araştırma sonuçları seçilen modele oldukça duyarlı olmaktadır. Bundan dolayı kayıp verilerle ilgili iyi tahminler yapılabilmesi için kayıp veri sürecinde yönelik iyi bir ön bilgiye sahip olmak gerekir (Allison ,2002).

Bu tür kayıpta kayıp veri yine sistematiktir ve verinin kayıp olma nedeni kayıp verinin kendisiyle ilişkilidir (Dural, 2010). Kayıp veriler varsayımsal veya hipotetik değişkenlerle alakalıdır. Örneğin; öğrencilere verilen bir okuma testinde, okuma güçlüğü çeken öğrencilerin bazı test sorularını boş bıraktığı görülebilir. Bu durumda okuma testindeki kayıp veriler öğrencilerin okuma güçlüğü ile ilişkilendirilebilir

(Baraldi ve Enders, 2010). X daima gözlenen, Y ise kayıp veri içeren değişken olmak üzere bu yapının formüsel olarak gösterimi şu şekildedir;

$$P(Y \text{ kayıp} / Y, X) = P(Y \text{ kayıp} / Y)$$

Buna göre Y değişkenindeki kayıp veri olasılığı X değişkenine bağlı değil ama Y değişkeninin kendisine bağlıdır (Allison, 2003; Dural, 2010; Schafer ve Graham, 2002).

Belirtilen kayıp veri mekanizmalarından TROK yapısında olanlar istatistiksel gücün azalmasına neden olsa da yansız parametre tahminleri (unbiased parameter estimates) yapılmasını sağlar. Yani, yapılan tahminler popülasyon değerlerine yakın çıkar. ROK yapısındaki kayıp veri mekanizması da yansız parametre tahminleri yapılmasını sağlarken, IE yapısındaki kayıp veri mekanizmalarının nedeni problemlidir olduğu için yanlış parametre tahminleri yapılmasına neden olmaktadır. Bu sebeplerden dolayı, TROK ve ROK mekanizmasındaki kayıp veriler ihmal edilebilirken, IE mekanizmasındaki kayıp veriler ihmal edilemez (Graham, 2009). Kayıp veri mekanizmasının ihmal edilebilir olması demek kayıp veri problemini önemsemek anlamına gelmemelidir. Bu durum, parametre tahminleri için kayıp veri mekanizmasının modellenmesine gerek kalmadan analizlerin yapılabileceği anlamına gelmektedir (Howell, 2007; McKnight ve diğerleri, 2007). Eğer, diğer değişkenlerin etkileri kontrol altına alındığında, bir değişkendeki kayıp veri içeren durumlar, aynı değişkendeki tam veri içeren durumlardan önemli derecede farklıysa ihmal edilemez kayıp vardır (Cheema, 2012).

#### *Kayıp verilerin rassallığının incelenmesi*

Kayıp veri mekanizmaları ile ilgili sınıflama kayıp verilerin rastgele dağılıp dağılmadığına göre yapılmaktadır (Osborne, 2013). Bu durumda kayıp verilerin rassallığının incelenmesi gerekir. Çünkü verinin matris üzerinde rassal olarak dağılması daha az problem ortaya çıkarırken, rassal olarak dağılmaması sonuçların genellenebilirliğini etkilemektedir (Tabachnick ve Fidell, 2013).

Kayıp verilerin rassallığının incelenmesinde 3 yöntem kullanılmaktadır. Bunlardan birincisi, veri seti kayıp veri içeren ve içermeyen durumlar olmak üzere ikiye ayrılır ve ilgilenilen değişkenler açısından iki grup arasında fark olup olmadığı incelenir. Bu durum incelenirken bağımsız gruplarda iki ortalama arasındaki farkın incelendiği t-testi uygulanır. Anlamli fark çıkmaması, kayıp verinin rassal dağılıma sahip olduğunu gösterir. İkinci yöntemde de değişkenler yine kayıp veri olan ve olmayan olmak üzere ikiye ayrılır ve aralarındaki Pearson korelasyon katsayısı hesaplanır. Buna göre düşük korelasyonlar her bir değişken çiftindeki rassallığı gösterir. Eğer tüm değişken çiftlerinde rassallık varsa veri yapısı TROK, bazı değişkenler arasında anlamlı korelasyonlar var ise veri yapısı ROK olarak sınıflandırılır. Üçüncü yöntem ise Little'ın TROK testi gibi tümel bir rassallık testinin uygulanmasıdır. Bu yöntem, sıklıkla kullanılan bir ki-kare testidir. Bu testte de anlamlı fark çıkmazsa veri yapısı TROK olarak sınıflandırılır. Eğer anlamlı fark çıkarsa rassal olmayan kayıp veri sürecini belirlemek amacıyla ilk iki yöntem kullanılabilir (Baygül, 2007; Kaspar, 2011).

Bu yöntemler içerisinde en sık kullanılanı Little tarafından geliştirilen TROK testidir. Bu yöntem 3 aşamada gerçekleştirilmektedir. İlk aşamada beklenti maksimizasyonu yöntemiyle tahmini olarak ortalama ve varyans-kovaryans matrisleri belirlenir. İkinci aşamada kayıp veri örüntüsüne göre gözlemler gruplandırılır ve her grup için gözlemlerin ortalaması alınır. Son aşamada ise gözlenen ve tahmin edilen ortalama değerlerin farkı alınarak, değerler varyans-kovaryans matrisi ve gruplardaki gözlem sayısına göre ağırlıklandırılır ve elde edilen istatistiğe göre varsayım test edilir (Yılmaz, 2014).

#### *Kayıp veri için uygun yöntemi seçme*

Kayıp veri mekanizmasının belirlenmesi, kayıp verilere uygulanacak uygun yöntemin seçilmesine, uygun yöntemin seçilmesi de analizler sonucunda doğru sonuca ulaşılmasını sağlar (Baygül, 2007; Buhi ve diğerleri, 2008). Bundan dolayı kayıp verilere uygulanacak yöntemi seçebilmek için bazı kriterler gerekmektedir. Allison (2009)'a göre bu kriterler şu şekildedir;

1. Yanlılıđı azaltma: Kayıp veriler parametre tahminlerinde yanlılıđa neden olduđu için seçilecek yöntem yanlılıđı olabildiđince düşük tutmalıdır.
2. Var olan bilgiyi en iyi şekilde kullanma: Kayıp veri olan durumlarda, seçilecek yöntem kayıp verileri analiz dıřı bırakmak yerine, parametre tahminlerinde elde olan bilgiyi en iyi şekilde kullanmalıdır.
3. Belirsizlik için iyi tahmin deđerleri verme: Standart hata, güven aralıđı ve  $p$  deđerleri için yapılan tahminler hassas ve dođru olmalıdır.

Kayıp veri mekanizmaları bütün veri setinin karakteristiđi olarak düşünülmemelidir. Bu mekanizmalar aslında kayıp verilere belirli analizleri yapabilmek için gerekli olan varsayımlardır. Sonuç olarak, hangi deđişkenin analize dahil edileceđine bađlı olarak, aynı veri seti TROK, ROK ve IE mekanizmalarına sahip olabilir (Baraldi ve Enders, 2010).

### 2.3. Kayıp Veri Yöntemleri

#### 2.3.1. Silme yöntemleri

Özellikle SAS ve SPSS gibi paket programların üzerinde durduđu en temel ve kolay yöntemlerdir (Baygöl, 2007). Liste ve çiftler bazında silme olmak üzere 2 silme yöntemi vardır.

1. Liste bazında silme-LBS (Listwise deletion): Liste bazında silme yöntemi herhangi bir deđişkende kayıp veri olması durumunda kayıp veri içeren birey ya da gözlemin analiz dıřı bırakılmasını kapsar (Cool, 2000; Howell, 2007; Imtiaz, 2007). Örneđin, 10 madde içeren bir ölçek bir öğrenci grubuna uygulanmış ve 5 numaralı katılımcı 7.soruyu boş bırakmış olsun. Çalışma kapsamında liste bazında silme yöntemi uygulanacaksa 5 numaralı katılımcı, analizlerden tamamen çıkartılarak gerekli işlemler uygulanır (Dural, 2010). Kayıp veri içeren durumlar çıkarıldığında, standart istatistiksel analizler

yapılabilmektedir. Söz konusu yöntem sadece kayıp veri mekanizmasının TROK olduğu durumda yansız sonuçlar üretmektedir (Allison, 2009; Baraldi ve Enders, 2010; Ginkel ve diğerleri, 2010; Peugh ve Enders, 2004). Çünkü son durumda elde edilen veri seti, orjinal veriden rastgele örneklem seçimine benzer şekilde elde edilmektedir ve rastgele örneklem seçimi yolu ile elde edilen verilerde yanlılık yoktur (Allison, 2009, Enders ve Bandalos, 2001, Ginkel ve diğerleri, 2010; Little ve Rubin, 1987).

Liste bazında silme yönteminin örneklem kaybının az olduğu durumda veya örneklemin büyük olduğu durumlarda kullanılması önerilmektedir çünkü bu durumlarda sonuçlar önemli derecede etkilenmemektedir (Acuna ve Rodriguez, 2004). Ancak, örneklem büyüklüğünün küçük olduğu durumda, kayıp veri mekanizması TROK olsa bile, yöntemin kullanılması istatistiksel gücün azalmasına, tip II hatanın olmasına (Buhi ve diğerleri, 2008; Cheung, 2007), daha büyük standart hatalara ve daha geniş güven aralığına neden olmaktadır (Allison, 2009, Cheung, 2007; Ginkel ve diğerleri, 2010). Çünkü bu durumda kısmen kullanılmayan veriler vardır (Graham, 2009). Liste bazında silme yönteminin en önemli dezavantajı örneklem kaybına neden olması olsa da tutarlı korelasyon ve kovaryans matrisi ürettiği için veriler modifikasyona gerek duyulmadan istatistiksel analizlerde kullanılabilir (Kim ve Curry, 1977). Örneklem büyükse (Acock, 2005) veya veri kaybı az ise (Graham, 2009) istatistiksel güç problem olmayacağı için liste bazında silme yöntemi kullanılabilir. Yöntemin diğer avantajları ise yapısal eşitlik modellemesinden, log-lineer analizlere kadar birçok istatistiksel analiz sürecinde kullanılabilir olması ve özel hesaplama yöntemleri gerektirmemesidir (Allison, 2002).

2. Çiftler bazında silme-ÇBS (Pairwise deletion): Erişilebilir hücre analizi olarak bilinen bu yöntem analiz sürecinde ulaşılabilen tüm verilerin kullanımına dayanmaktadır. Yöntem, kovaryans veya korelasyon matrisi ile açıklanmaktadır. Elde olan veri veya veri çiftindeki tam veriler kullanılarak korelasyon ve kovaryans ölçüleri ile ortalama, standart sapma vb. tanımlayıcı

istatistikler hesaplanır (Allinson, 2002; Allison, 2009; Peugh ve Enders,2004). Yani yöntem, kayıp veriyi tahmin sürecinde ham veriler yerine ortalama, varyans, kovaryans vb. birinci ve ikinci momentleri kullanır (Allison, 2003). Veriler, veri çiftine göre hesaplamaya dâhil edilir veya hesaplamadan silinir. Örneğin, bir katılımcının  $Y, X_1, X_2, X_3$  değişkenlerindeki verileri tam,  $X_4$  ve  $X_5$  değişkenlerinde de eksik verileri olsun. Bu durumda  $r_{YX1}, r_{YX2}, r_{YX3}$  değişkenlerindeki korelasyon değerleri hesaplanırken  $r_{yx4}, r_{yx5}$  korelasyon değerleri hesaplanamaz yani kovaryans veya korelasyon matrisine dahil edilmez (Howell, 2007). Yöntemin liste bazında silme yönteminden farkı, kayıp veri içeren bireyin tamamını listeden çıkarmaktansa sadece kayıp veri içeren durumları kaldırmasıdır (Cheema, 2012). Bu durumda da istatistiksel güç kaybı daha az olmaktadır (Ginkel ve diğerleri, 2010).

Yöntem, doğrusal modellerde (Doğrusal regresyon, faktör analizi, yapısal eşitlik modellemesi vb.) liste bazında silme yönteminin alternatifi olarak düşünülmektedir. Kayıp veriyi kestirim sürecinde, daha fazla veriyi dikkate aldığı için daha az örneklem değişkenliğine (sampling variability) neden olmakta ve bundan dolayı standart hataları daha küçük ve doğru bir şekilde hesaplamaktadır. Bundan dolayı liste bazında silme yöntemine göre daha iyi bir yöntem olduğu ifade edilebilir. Ancak bunun yanı sıra bazı sınırlılıkları vardır. Yöntemin uygulanabilmesi için verinin TROK yapısına uygun olması gerekmektedir. Aksi takdirde parametre tahminleri yanlı olmaktadır (Allison, 2002; Allison, 2009; Baraldi ve Enders, 2010; Enders, 2010; Ginkel ve diğerleri, 2010). Diğer bir sınırlılığı ise örneklem büyüklüğüne hassas olmasıdır. Örneklem büyüklüğü yeterli düzeyde değilse kovaryans matrisi pozitif tanımlı olmaktan çıkabilmekte, parametre tahminleri için dönüştürülememekte ve standart hatalar doğru bir şekilde hesaplanamamaktadır (Allison, 2002; Allison 2009). Başka bir ifade ile iki değişken arasında kabul edilebilir olmayan yani negatif bir korelasyon değeri elde edilebilmektedir (Cheung, 2007; Graham, 2009). Bunun nedeni ise; her korelasyon değeri, herhangi iki değişkenin de verisinin olduğu durumlara göre hesaplanır. Kayıp veriden dolayı her seferinde farklı bir alt örnekleme

göre korelasyon değeri hesaplanacağı için örneklem sayısı her seferinde farklı olmakta ve bundan dolayı parametre tahminleri yanlı olabilmektedir (Allison, 2009; Ginkel ve diğerleri, 2010; Graham, 2009; Howell, 2007).

Liste bazında ve çiftler bazında silme yöntemleri genellikle istatistiksel gücün azalmasına ve yanlı sonuçlar elde edilmesine neden olmaktadır. Bundan dolayı araştırmacıların bu analizleri tercih ettiklerinde dikkatli olması ve TROK mekanizmasının sağlandığını görmesi gerekmektedir (Ginkel ve diğerleri, 2010).

### 2.3.2. Atama yöntemleri

Kayıp veriyi kestirim sürecinde kullanılan bir diğer yöntem de atama yöntemleridir. Bu yöntemlerde tam veri içeren değişken veya durumlar kullanılarak kayıp veriler yerine atama yapılır (Kaspar, 2011). Kayıp veriler için atama yöntemlerinin tercih edilmesinin nedeni; istatistiksel hesaplamaları kolaylaştırmak, doğru parametre tahminleri yapmak ve birçok değişken için olabildiğince çok durumu elde tutmaktır (Kim ve Curry, 1997). Atama yöntemleri aslında bir tahmin sürecidir ve bundan dolayı;

- Atamalar kayıp verilerin tahmini dağılımına göre yapılmalıdır,
- Bütün gözlemlenen veriler atama sürecinde kullanılmalıdır,
- Tahmin için kullanılacak model, ataması yapılacak değişkenlerle bağlantılı olmalıdır,
- Tahmin için kullanılacak model, veri setinin ranjı dışında tahmin değerleri üretmemelidir,
- Tamamlanmış veri setlerindeki dağılımı korumak için atamalar ortalamalara göre değil, kayıp verilerin tahmini dağılımına göre yapılmalıdır, yani artık veya rastgele hata değerleri ortalamalara eklenmelidir (Little, 1988).

Atama yöntemleri şu şekilde sınıflanmaktadır;

1. Yaklaşık Değer Atama (Tahmin) Yöntemleri: IBM (2012) ve Kalaycı (2010) kayıp veriler için yaklaşık değer atama yöntemlerini şu şekilde tanımlamıştır;

- a. Seriler ortalaması-SO (Series mean): Kayıp veriler yerine tüm serinin ortalamasını atar.
  - b. Yakın noktaların ortalaması-YNO (Mean of nearby points): Kayıp verilerin altında ve üstünde bulunan tam verilerin ortalamasını alarak kayıp verilerin yerine atar.
  - c. Yakın noktaların medyanı-YNM (Median of nearby points): Kayıp verilerin altında ve üstünde bulunan tam veriler kullanılarak medyan değeri hesaplanır ve kayıp verilerin yerine atanır.
  - d. Doğrusal değer kestirimi-DDK (Linear interpolation): Kayıp veriden önceki son tam gözlem ile kayıp veriden sonraki ilk tam gözlemi, kayıp veri değerini kestirmede kullanır. Serideki ilk ve son gözlem kayıp veri içeriyorsa, kayıp veri yerine değer yerleştirilemez.
  - e. Noktanın doğrusal eğimi-NDE (Linear trend at point): Mevcut seriler 1'den n'e kadar ölçeklendirilmiş bir endeks değişkeni üzerinde dönüştürülür ve eksik veriler tahmini değerlere göre yerleştirilir. Başka bir ifade ile mevcut verinin sahip olduğu eğilimle uyumlu değer atanır.
2. Ortalama atama- OA (Mean imputation): Yöntemde, kayıp veriler yerine mevcut verilerin ortalaması atanmaktadır (Baraldi ve Enders, 2010; Brown ve Kros, 2003, Buhi ve diğerleri, 2008; Howell, 2007). Kayıp veriler yerine atama yaparken, tam verilerin ortalamasını atadığı için, atama yöntemleri içerisinde en kolay yöntem olarak belirtilmektedir. Teorik olarak kayıp verilerle ilgili herhangi bir bilgi yoksa ve kayıp veriler rastgele dağılıma sahipse ortalama atama yöntemi en iyi tahmin değerini vermektedir (Brown ve Kros, 2003; Gelman ve Hill, 2012; Kang, 2013; Osborne, 2013). Ancak, kayıp verilerin rastgele dağılıma sahip olmadığı durumda yapılan ortalama atama işlemi yanlış sonuçlara neden olmaktadır (Kang, 2013). Atama işlemi sonucunda tam ve atanmış veriler için elde edilen ortalama, tam verilerle elde edilen ortalamaya eşit çıkacaktır (Bal, 2003). Yöntem araştırmacılar tarafından sıklıkla tercih edilen bir yöntem olsa da, yöntemin uygulanacağı verilerin geçerliği kanıtlanmış ve normal dağılıma sahip veriler olmasına dikkat edilmelidir (Brown ve Kros, 2003).

Ortalama atama yöntemi veri yapısı TROK olduğunda kayıp veriyi tahmin konusunda geçerli veriler üretmesine rağmen varyans ve kovaryans kestiriminde yanlı tahminlerde bulunmaktadır. Çünkü kayıp veriler yerine aynı ortalamayı atadığı için kayıp veriler arasındaki değişkenliği sınırlamaktadır (Bennett, 2001; Lakshminarayan, Harp ve Samad, 1999; Sinharay ve diğerleri, 2001; Tabachnick ve Fidell, 2007). Bu durum standart hataların küçülmesine, varyansın olduğundan küçük hesaplanmasına, kovaryansların negatif eğilimli çıkmasına neden olmakla birlikte atama sonrası elde edilen yeni veri setinin toplumu yeterince yansıtmamasına neden olabilir (Baygöl, 2007; Brown ve Kros, 2003; Little ve Rubin, 1987). Bundan dolayı yöntemin, kayıp veri oranı az ve ranjı düşük veri setlerinde kullanılmasının daha uygun olduğu belirtilmiştir (Karal, 2014). Ancak kayıp olmayan tüm verileri kullanması, hesaplama kolaylığından dolayı özel yazılım programları gerektirmemesi, (Baygöl, 2007; Brown ve Kros, 2003) kayıp verileri gerçek bir değerle değiştirdiği için örneklem büyüklüğünü ve buna bağlı olarak istatistiksel gücü koruması açısından avantajlı bir yöntemdir (Bal, 2003; Buhi ve diğerleri, 2008; Cool, 2000).

3. Regresyon ataması- RA (Regression imputation): Koşullu ortalama değer atama (conditional mean imputation) olarak da adlandırılan yöntemde amaç, bağımsız değişkenler aracılığıyla bağımlı değişken değerlerinin test edilmesidir. Kayıp veriler bağımlı değişken, diğer veriler de bağımsız değişken kabul edilerek, kayıp veriler için tahminde bulunulur. Tahminde bulunurken tam veri seti içeren durumlar kullanılarak regresyon denklemi yazılır ve bu denklem yoluyla kayıp veriler yerine atama yapılır (Baraldi ve Enders, 2010; Brown ve Kros, 2003; Cool, 2000; Downey ve King, 1998; Dural, 2010; Tabachnick ve Fidell, 2007). Örneğin, üniversite öğrencilerinin matematik dersindeki başarıları (X) ve ölçme değerlendirmeye yönelik tutumlarının (Y) incelendiği bir çalışma olsun. 50 kişilik öğrenci grubundan oluşan bu çalışmada matematik dersindeki başarı değişkeninde eksik veri yer almazken, ölçme ve değerlendirmeye yönelik tutum değişkeninde kayıp veri

bulunsun. Regresyon ataması yönteminde öncelikle başarı değişkeninden, tutum değişkeninin yordandığı bir eşitlik elde edilmektedir;

$$Y=a+bX$$

Sonra tutum değişkeninin boş bir hücrelerine denk gelen başarı değişkeni, bu denklem üzerine yerleştirilir ve tutum değişkeninin değeri hesaplanır. Bu regresyon eşitliği, tutum değişkeninde kayıp veri içeren tüm durumlar için uygulanır. Sonrasında 2 değişken arasındaki ilişkiyi incelemek için korelasyon katsayısı hesaplanır (Dural, 2010). Regresyon ataması yöntemi kayıp veri yerine tek seferde atama yaptığı için tekli atama yöntemi (ortalama atama, hot deck atama gibi) olarak da adlandırılmaktadır (Peng, Harwell, Liou ve Elman, 2006).

Bu yöntem uygulanırken dikkat edilmesi gereken iki nokta vardır; bunlardan birincisi değişkenler arasındaki ilişkinin yeterli düzeyde olması, diğeri ise bağımlı değişkenin bağımsız değişken(ler) tarafından iyi bir şekilde açıklanmasıdır. Eğer bu durum sağlanmıyorsa ortalama atama yönteminin kullanılmasının daha uygun olduğu belirtilmektedir (Bal, 2003). Bunun yanı sıra kayıp veri içeren değişkenlerin her biri için ayrı bir regresyon işlemi uygulanmalı ve veri yapısı TROK veya ROK olmalıdır (McKnight ve diğeri, 2007). Çünkü veri yapısının TROK ve ROK olduğu durumlarda yansız parametre tahminleri elde edilebilmektedir (Baraldi ve Enders, 2010).

Regresyon atamasının bir avantajı, atama yapılacak değişkende bulunan her bir kayıp veri için farklı bağımsız değişken kümesi kullanmasıdır. Ortalama atama yöntemine göre avantajı ise kayıp değer içeren değişkenler için varyans ve kovaryansı korumasıdır. Çünkü bir değişkende yer alan her bir kayıp durumu diğer değişkenlere bağlı olduğu için her seferinde farklı bir tahmin değeri elde edilmektedir (Oğuzlar, 2001). Silme yöntemlerine göre avantajı ise elde daha çok veriyi tutmasıdır (Roth, 1994). Yöntem çok fazla

hesaplama işlemi gerektirdiği için pratik bir yöntem olarak görülmesi de SPSS gibi programlarla birlikte uygulama kolaylığı da gelmiştir (Cool, 2000).

Regresyon atamasının dezavantajları ise şu şekildedir; kayıp veriler diğer verilerden tahmin edildiği için veri içinde var olan ilişki daha da kuvvetlenmiş olur ve değişkenlik azalır (Brown ve Kros, 2003; Newman, 2003). Kayıp veriyi tahmin etmede kullanılan bağımsız değişkenler, kayıp verinin iyi bir tahmin edicisi değilse sonuç ortalama atama ile benzer olacaktır. Bunun yanı sıra tahmini yapılan veriler için herhangi bir kısıtlama olmadığı için, yapılan tahminler o değişken için belirlenen aralığın dışında çıkabilir (Brown ve Kros, 2003; Peng ve diğerleri, 2006; Tabachnick ve Fidell, 2007). Örneğin, bağımlı değişken değerleri, 0-10 arasında değişen değerler alıyorsa, kayıp veri değeri 12 olarak atanabilir. Bundan dolayı bazı düzeltme işlemleri gerekebilir (Brown ve Kros, 2003). Bunlara ek olarak değişkenlik azaldığı için regresyon katsayılarının standart hataları olduğundan az hesaplanır. Ancak SPSS paket programında yöntem uygulanırken her atama yapılan veri için hata terimi eklenir. Bu da standart hatalardaki negatif yanlılığı kısmen azaltmaktadır (Howell, 2007). Hata terimi modele dâhil edilmezse varyasyon olduğundan küçük çıkar (Baygül, 2007). Hata teriminin eklendiği regresyon ataması stokastik regresyon ataması (stochastic regression imputation) olarak adlandırılmaktadır ve normal regresyon atamasına göre tam veri setine daha yakın sonuçlar vermektedir (Baraldi ve Enders, 2010; Ginkel ve diğerleri, 2010). Yanlılığın azalması için hata teriminin eklenmesinin yanı sıra doğrusallık varsayımının sağlanması gerekir. Yani, tüm değişkenler birbirleri ile doğrusal olarak ilişkili olmalıdır (Hertel, 1976; Young, Weckmen ve Holland, 2011).

4. En çok olabilirlik- EÇO (maximum likelihood): Yaklaşımın temel amacı kayıp veri tahmini için gözlemlerin olasılığını maksimum yapacak verilerin seçilmesidir (Cheema, 2012). Kayıp veri kestiriminde kullanılan bir yöntem olmaktan çok parametre kestirimi için kullanılan olasılık temelli bir yaklaşımdır (Demir ve Parlak, 2012). Bu yaklaşım daha çok yapısal eşitlik

modellemesinde, çoklu normallik varsayımı altında, tahmin için kullanılmaktadır ve özellikle ROK varsayımının sağlandığı durumlarda daha iyi tahmin değerleri vermektedir (Allison, 2003; Baraldi ve Enders, 2010). ROK mekanizmasının yanı sıra TROK mekanizmasında da yansız parametre tahminleri yapmaktadır (Enders ve Bandalos, 2001). Tahmin sürecinde, üretilen örnek bir veri üzerinden populasyon parametreleri tahmin edilir. Kayıp verileri direk olarak doldurmaktansa hem tam hem de kayıp verileri kullanarak parametre tahmini yapar (Baraldi ve Enders, 2010; Collins, Schafer ve Kam, 2001).

Kayıp verileri tahmin sürecinde beklenti maksimizasyonu algoritması, çoklu atama ve tam bilgi en çok olabilirlik yöntemlerinde en çok olabilirlik yöntemi temelli tahminler yapılmaktadır (Cheung, 2007). En çok olabilirlik süreci tekrarlama işlemini gerektirir. Modelleme için kullanılan program, ilgili parametrelerin tahmini için farklı değerler dener ve böylece örnek veri seti için en uygun değerler belirlenir. Bir grup parametre değişkeni için verilerin ayarlanması log-likelihood değeri ile belirlenir. Bu değer, populasyon verilerinin normal dağılması şartı ile bir örnekleme bulunan kayıp verilerin olasılık değerlerinin belirlenmesini sağlar. Log-likelihood değeri her bir tekrar sürecinde hesaplanır (Peugh ve Enders, 2004). Bu değer, gözlemlenen veri noktaları ile ilgili parametreler (örneğin, ortalama) arasındaki uzaklığın belirlendiği değerdir ve amaç bu iki değer arasındaki uzaklığı minimize ederek parametre değerlerini belirlemektir. Bu durum en küçük kareler yöntemi ile benzerdir. Çünkü bilindiği üzere, bu yöntemde de veri noktaları ile bağımlı değişken arasındaki uzaklık minimize edilerek regresyon katsayılarını tanımlanmaktadır (Baraldi ve Enders, 2010).

Yöntemin uygulanabilmesi için ilgili bütün değişkenlerin ortak dağılımının modellenmesi ve olasılığı maksimize etmek için sayısal bir yöntem gereklidir. Örneğin tüm değişkenler kategorik ise multinominal veya log-linear modeller kullanılırken; değişkenler sürekli ise çoklu normallik varsayımı altında bir model kullanılır. Bu varsayımda her değişkenin normal

dağıldığı ve diğer değişkenlerin doğrusal fonksiyonu olarak açıklandığı belirtilir. Çoklu normallik varsayımı altında tahminler beklenti maksimizasyonu algoritması veya doğrudan beklenti maksimizasyonu (Direct maksimum likelihood) ile yapılmaktadır. Bu yöntemlerden doğrudan beklenti maksimizasyonu yöntemi standart hata kestirimlerini daha doğru vermektedir ve fazla tanımlanmış modellerde kullanımı daha uygundur. Ancak birçok yazılım programında beklenti maksimizasyonu yöntemi tanımlı olduğu için, bu model daha sık kullanılmaktadır (Allison, 2009). Allison (2009) en çok olabilirlik yönteminde beklenti maksimizasyonu algoritması ile tahmin sürecini şu şekilde açıklamıştır;

1. Liste bazında silme veya çiftler bazında silme yöntemi ile ortalama, varyans ve kovaryans değerleri hesaplanır.
2. Gözlemlenen değişkenler kullanılarak regresyon denklemi ile kayıp değişkenler tahmin edilir. Regresyon parametreleri ilk adımda hesaplanan ortalama, varyans ve kovaryans değerleri ile hesaplanır.
3. Regresyon denklemi kullanılarak kayıp veri içeren değişkenlerin değerleri tahmin edilir.
4. Bütün gerçek ve atanan veriler kullanılarak ortalama, varyans ve kovaryans değerleri tekrardan hesaplanır.
5. Yakınsaklık sağlanana kadar 2.adıma tekrar dönlür.

Gerekli varsayımlar sağlandığı sürece en çok olabilirlik yöntemi tutarlı, asimptotik normal ve asimptotik etkili tahmin değerleri üretmektedir (Allison 2002; Allison 2009; Agresti ve Finlay, 1997). Tutarlıktan kasıt yapılan tahminlerin büyük örneklem için yansız olmasıdır. Yani, örneklem büyüklük bakımından evrene yaklaştıkça yapılan tahminlerin evren parametresine yaklaşma olasılığı da artar (Baykul ve Güzeller, 2013). Asimptotik etkililikten kasıt standart hataların küçük hesaplanmasından dolayı yapılan tahminlerin etkili olmasıdır. Asimptotik normallik ise tekrarlı yapılan örneklemede tahminlerin yaklaşık olarak normal dağılmasıdır. Burada yine örneklem büyüklüğü arttıkça tahminlerin normale yaklaşma olasılığı da artmaktadır (Allison, 2002; Allison, 2009). Burada “asimptotik” ifadesi ile

anlatılmak istenen, belirtilen durumların yaklaşık olarak doğru olduğudur ve belirtilen durumların sağlanması olasılığı, örneklem büyüklüğünün artması ile artmaktadır (Allison, 2002).

Yöntemin bazı olumsuz yanları vardır; bunlardan biri özel yazılım programları gerektirmesidir. Bir diğeri de kayıp veri içeren bütün değişkenlerin ortak dağılımını gösteren parametrik bir model gerektirmesidir. Bu tür modellerin tasarlanması zor olmakla birlikte, sonuçlar model seçimi için hassas olabilmektedir (Allison, 2009). Ancak, geliştirilen yeni yazılım programlarıyla bu tür zorluklar minimize edilerek kullanıcılara uygulama kolaylığı sağlanmıştır (Collins ve diğerleri, 2001). Yöntemin uygulaması AMOS gibi yapısal eşitlik modellemesi programlarında yapılabilmektedir (Enders ve Bandalos, 2001). Fakat istatistik yazılımlarının kullanımlarındaki en önemli sorun gerek çoklu atama gerekse en çok olabilirlik yöntemlerinde atamalar seçkisiz olarak yapıldığından dolayı her bir yazılım farklı sonuç üretmektedir. Bunların yanı sıra, en çok olabilirlik tahminleri küçük örneklerde zayıf çalışmaktadır (Allison, 2009).

5. Beklenti maksimizasyon algoritması- BM (Expectation maximization): Dempster, Laird ve Rubin (1977) tarafından önerilen bir yöntemdir. Beklenti maksimizasyonu bir en çok olabilirlik yöntemidir (Graham, 2009; Jing, 2012; Karal, 2014; McKnight ve diğerleri, 2007) ve bütün kayıp verilerin en çok olabilir değerleriyle atanması sonucu yeni bir veri seti oluşturur (Newman, 2003). Yaklaşım, bütün gözlemlenen değerlerin arasındaki ilişkiye dayanır. Kayıp veri değerleri tekrarlı atamalar sonucunda belirlenir. Ardarda yapılan her tekrarlı atama sonucunda daha fazla bilgi edinilir çünkü her tekrar, önceki tekrarlarda yer alan bilgileri kullanmaktadır. Tekrarlama işlemi, kovaryans matrisinden önce ve sonra yapılan atamalarda benzer sonuçlar elde edilene kadar devam ettirilir. Bu süreç kısa sürmekle birlikte eğer kayıp veri miktarı fazla ise biraz zaman almaktadır (Acock, 2005). Yöntemde, kayıp veriler yerine direk atama yapılmaz. Bunun yerine öncelikle kovaryans matrisinin en çok olabilirlik değerleri elde edilir ve bu kovaryans matrisine göre ilerleyen

analizlerde kayıp veriler yerine atama yapılır. Yani, yöntemde parametre tahmini için ek bir veri hazırlama aşaması gerekmektedir (Enders, 2001 (b)).

Beklenti maksimizasyonu iki adımdan oluşmaktadır; Beklenti adımı- B (expectation) ve maksimizasyon adımı- M (maximization). B adımında kayıp veriler, beklenen değerlerle tamamlanır, M adımında ise B adımında tahmin edilen değerler kullanılarak, tam veri seti üzerinden parametre tahminleri yapılır. B adımında öncelikle var olan verilerle parametre değerleri (ortalama, kovaryans gibi) hesaplanır. Sonra bu başlangıç verilerine göre regresyon denklemi kullanılarak kayıp veri ataması yapılır. B adımı aslında beklenen değerleri elde etmek için ard arda gerçekleştirilen regresyon ataması gibidir ve regresyon ataması yapılırken hata terimi de eklenmektedir. Bu aşama tamamlandıktan sonra, maksimizasyon adımında, ataması yapılan veri seti ile orijinal veri seti kullanılarak parametrelerin yeni değerleri hesaplanır (Cheung, 2007; Enders, 2001 (b); Newman, 2003; Peugh ve Enders, 2004; Schlomer, Bauman ve Card, 2010). M adımında, B adımında tahmin edilen kayıp veri değerlerinin maksimizasyonu yapılır. Maksimizasyon aşamasında elde edilen yeni kovaryans matrisi ve regresyon katsayıları, beklenti aşamasında yeni tahminler yapmak amacıyla tekrar kullanılır (Duncan ve diğerleri, 1998, Enders, 2001(b)). Bu süreçte tekrarlama işlemi uygulanır ve işlem, başta belirlenen yakınsaklaşma kriterine ulaşılan kadar (örneğin; iki ardışık parametre tahmini birbirine yakın değerler verene kadar) devam ettirilir (Allison, 2003; Dempster ve diğerleri, 1977; Graham, 2009). Yani, bu işlem ilk adımdaki kayıp veri matrisi ile 2.adımda oluşturulan veri matrisi arasında anlamlı fark kalmayıncaya kadar devam ettilir (Baygöl, 2007). Sonuç olarak, kovaryans matrisi ve ortalama vektörleri için en çok olabilirlik tahminleri yapılmış olur (Peugh ve Enders, 2004). Süreç doğrusallık ve normallik varsayımına dayanır (Baygöl, 2007). Allison (2009)'a göre çoklu normallik varsayımı altında beklenti maksimizasyonu adımları şu şekildedir;

1. Modelde tahmin edilen parametreler ortalama, varyans ve kovaryans olduğu için öncelikle liste bazında veya çiftler bazında silme yöntemi ile bu değerler elde edilir.
2. Her kayıp veri mekanizması için regresyon denklemi yazılarak gözlemlenen veriler yardımıyla kayıp veri değerleri tahmin edilir. Kayıp veri değerleri tahmin edilirken tam verilerle birlikte ilk adımda belirlenen ortalama, varyans ve kovaryans değerleri de kullanılır.
3. Regresyon denklemleri bütün kayıp veri içeren değişken ve durumlar için tahmini değerler üretmede kullanılır.
4. Bütün gerçek ve atama yapılan veriler kullanılarak ortalama, varyans ve kovaryanslar yeniden hesaplanır. Ortalamalar için standart formül iyi çalışırken, atanan verilerde yanlılığı azaltmak için varyans ve bazen de kovaryanslar için düzeltme faktörü eklenmektedir.
5. 2. adıma dönülerek yakınsaklık sağlanana kadar işleme devam edilir.

Beklenti maksimizasyonunda temel çıktılar ortalama, varyans ve kovaryansların maksimum olabilirlik kestirimleridir. Bu çıktılar aynı zamanda kestirim sürecinin bir parçası olarak da kullanılmaktadır. Bu yöntemin bir avantajı belirtilen temel çıktı tahminlerini hemen hemen yansız bir şekilde yapmasıdır (Howell, 2007). Özellikle veri yapısı ihmal edilebilir olduğu zaman yani veri yapısı TROK veya ROK olduğunda yansız parametre tahminleri yapmaktadır (Dragset, 2009). Diğer bir avantajı ise kullanımın kolay olması ve birçok paket programla gerçekleştirilebilmesidir. Çoklu normallik varsayımı altında beklenti maksimizasyonunu gerçekleştiren birçok paket program vardır. Bu programlar; SPSS, BMDP, SAS, S-PLUS ve SYSTAT'tır (Allison, 2003).

Yöntemin en önemli dezavantajı, atama sürecinde değişkenliği koruyamamasıdır. Bunun iki nedeni vardır; bunlardan birincisi regresyon atamasıdır, çünkü atanan veriler regresyon doğrusu boyunca düşeceği için değişkenlik azalacaktır. Diğer neden ise değişkenliğin kovaryans matrisinden dolayı bozulmasıdır, çünkü veriler hata bileşeni olmadan atanmaktadır. Bunun sonucunda da ilerleyen analizler için standart hatalar negatif eğilimli olmaktadır

(Enders, 2001(b)). Yani, lineer model programları ile üretilen standart hata ve test istatistiklerini doğru hesaplanamamakta ve gereğinden fazla tanımlanan modellerde yapılan tahminler, kovaryans matrisindeki sınırlılıklardan dolayı, etkili olamamaktadır (Allison, 2002). Bundan dolayı, beklenti aşamasında kovaryans matrisindeki negatif yöndeki yanlılığı düzeltmek amacıyla her atama yapılan veri için düzeltme faktörü veya hata terimi eklenmektedir (Enders, 2001(b); Graham, Hofer ve Piccinin, 1994; Howell, 2007). Bunların yanı sıra çok fazla kayıp veri olduğunda yakınsaklaşma süreci uzun zaman almaktadır (Bennett, 2001; Dragset, 2009; Imtiaz, 2007).

6. Çoklu atama- ÇA (Multiple imputation): Yöntem, Rubin (1987) tarafından önerilmiş olup BM yönteminin tekrarlı bir uzantısıdır (Olinsky, Chen ve Harlow, 2003). Her kayıp veriyi tek bir değer ile tamamlamak yerine, kabul edilebilir iki ya da daha fazla verinin atanmasını sağlar. BM yöntemi gibi atama yöntemlerinin silme yöntemleri gibi geleneksel yöntemlere göre avantajları varken, çoklu atama yönteminin de BM yöntemi gibi tekli atama yöntemlerine göre üstün olduğu yönler vardır. Özellikle tekli atamalarda, çoklu atama verilerinde olduğu gibi atama yapılan veriler arasındaki farklar görülememektedir, çünkü tek bir atama verisi elde edilmektedir. Bundan dolayı standart hata kestirimleri olduğundan az hesaplanmaktadır (Acock, 2005).

Çoklu atamada her bir kayıp veri için  $m$  tane veri ataması yapılır ve  $m$  tane atama sonucunda  $m$  adet tam veri seti elde edilir. Sonrasında her veri seti tek bir veri seti altında birleştirilerek (örneğin, ortalamaları alınarak), atanan veriler gerçek verilermiş gibi standart tam veri seti sürecine göre analiz edilir (Ginkel ve diğerleri, 2010; Granberg- Rademacker, 2007; Rubin, 1987). Bu yöntemde 2-10 arası tamamlanmış veri setinin kullanımının ideal olduğu belirtilmektedir (Cheema, 2012; Rubin, 1987). Schafer (1997) ve Graham (2009)'a göre de 3, 4 veya 5 tamamlanmış veri setinin kullanımının ideal olduğunu belirtmiştir. Çünkü bu civarda kullanılan veri setlerinde çoklu atama aralığının, daha fazla değişkenlik gösterdiği gözlemlenirken, 10'dan fazla kullanılan veri setlerinde değişkenliğin bir süre sonra önemli derecede azaldığı belirtilmiştir. 3 veya 5 adet

veri seti kullanıldığında değişkenliğin fazla olması sonuçların güvenilir olmadığı anlamına gelmemelidir. Çünkü değişkenlik, belirsizliğin bir parçası olan hata faktörünü içermektedir (Schafer, 1997). Yöntem, her atama için farklı bir çözüm önerisi üretir. Eğer m tane atama sonucu birbirine benzer çıkarsa, yöntem desteklenmeli. Eğer sonuçlar önemli derecede farklı çıkarsa, bu farklılık standart hatalarla ilişkilendirilmelidir (Acock, 2005).

Çoklu atama yöntemi 3 adımda uygulanmaktadır. Bu adımlar; atfetme (imputation), analiz etme (analysis) ve bir araya getirme (pooling) dir. Belirtilen bu adımların açıklaması şu şekildedir;

- İlk adım atfetme evresidir ve veri çoğaltma yolu ile 5-10 arasında (veya 2-10 arası) veri seti üretilir (Arnold ve Kronmal, 2002).
- İkinci adım analiz evresidir. Bu evre atama evresinden daha basittir çünkü eksik veri problemi ortadan kalkmıştır (Bal, 2003). Bu aşamada veri çoğaltma yapılarak her veri seti için ayrı ayrı model tahmini yapılır (regresyon, lojistik regresyon, YEM gibi). Veri çoğaltma işlemi Markov Chain Monte Carlo (MCMC) yöntemi ile yapılmaktadır (Howell, 2007; Rose ve Fraser, 2008). MCMC yöntemi dışında çoklu atama temelinde geliştirilmiş “tanımlanmış atama (deterministic imputation)” ve “seçkisiz atama (random imputation)” yöntemleri de vardır (Allison, 2009).
- Üçüncü adım birleştirme evresidir ve her veri seti için parametre tahminleri (faktör yükleri, korelasyon değeri, regresyon katsayısı gibi) ve standart hatalar hesaplanmaktadır. Sonuçlar, çoklu analizlerle parametre tahminlerinin ortalamaları alınarak hesaplanır (Acock, 2005; Arnold ve Kronmal, 2002; Baraldi ve Enders, 2010; Schlomer ve diğerleri, 2010).

Çoklu atama yöntemiyle elde edilen tekli atamaların parametre tahminleri ve standart hataları şu formüllerle birleştirilir (Baygül, 2007; Cheema, 2012; Rubin, 1987);

$$\bar{q} = \frac{1}{m} \sum_{k=1}^m \hat{q}_i$$

$m$ = Ataması yapılmış ve analiz edilmiş küme sayısı

$\bar{Q}_i$  = Analiz edilmiş i.kümeden tahmin

Formülden görüleceği gibi çoklu atamalardan elde edilen nokta tahmini her analizden elde edilen sonucun ortalamasıdır.

Standart hataların birleştirilmesi, nokta tahminine göre biraz daha karışıktır. Standart hatalar gruplar içi ve gruplar arası varyasyonun birleştirilmesi ile elde edilir. İşlem adımları şu şekildedir;

1. Öncelikle her veri setinden elde edilen standart hataların karesi alınır ve yapılan  $m$  tane replikasyona göre ortalaması alınır.
2. Replikasyonlara göre parametre tahminlerinin varyansları hesaplanır.
3. 1.ve 2.adımda elde edilen veriler birleştirilerek bu verilerin karekökü alınır (Allison, 2003; Allison, 2009).

Rubin (1987)'e göre standart hata hesabının formülle ifadesi şu şekildedir;

$$\sqrt{\frac{1}{m} \sum_{k=1}^m S_k^2 + \left(1 + \frac{1}{m}\right) \left(\frac{1}{m-1}\right) \sum_{k=1}^m (b_k - \bar{b})^2}$$

$m$ = Ataması yapılmış ve analiz edilmiş küme sayısı

$S_k$ = k.veri setindeki standart hata

$b_k$ = k.veri setinden elde edilen parametre tahmini

En çok olabilirlik yönteminin aksine, çoklu atama her türlü veri ve modele uygulanabilmektedir. Ancak, çoklu atama her seferinde seçkisiz atamadan dolayı farklı tahminler üreteceği için aynı veri setinden, aynı metotları kullanarak farklı sonuçlar elde edilebilir (Allison, 2003; Allison,2009). Bunun yanı sıra korelasyon değerini olduğundan fazla gösterebilir. Örneğin X ve Y iki değişken olsun ve X değişkeninde bazı veriler rastgele kayıp olsun. X değişkenine çoklu atama ile atanan veriler sonucunda Y değişkeni ile aralarında 1.00 gibi mükemmel bir korelasyon değeri çıkabilir. Bu sorun, atanan veriler arasında seçkisiz değişkenlik (Random variation) sağlanarak çözülebilir

(Allison, 2003). Yöntemin avantajı parametre tahminlerinin standart hataları; her veri setinin standart hataları ve parametre tahminlerinin veri setindeki saçılımı dikkate alınarak hesaplanır. Atama yapılmış verilerden elde edilen bu standart hatalar hipotez testi ve/veya parametre tahminlerinde güven aralığının belirlenmesinde kullanılır. Her atama sonunda seçkisiz değişkenlikten dolayı farklı sonuçlar elde edildiği için standart hatalar daha doğru hesaplanır ve böylece sonuçlara dair daha doğru çıkarımlar yapılır (Schlomer ve diğerleri, 2010). Standart hatalardaki yanlılığı azaltmasından dolayı çoklu atama yöntemi, tekli atama yöntemlerine göre daha avantajlıdır. Bunların yanı sıra liste bazında veya çiftler bazında silme yöntemleri sadece TROK varsayımı altında geçerli çıkarımlar yaparken, çoklu atama hem TROK hem ROK varsayımı altında geçerli çıkarımlar yapar (Ginkel ve diğerleri, 2010). Rubin (1987) ise tekli atama yöntemlerine göre avantajlarını şu şekilde özetlemiştir;

- Çoklu atama rastgele yapıldığında tahmin değerinin etkililiğini artırır.
- Yanıtsız modeller (model for nonresponse) için birden fazla atama yaptığı için yapılan atamalar birleştirilerek geçerli çıkarımlar elde edilir.
- Birden fazla model içerisinde tekrarlı atamalar yaptığı için birden fazla tam veri seti elde edilir bundan dolayı hassas çıkarımlar yapılmasını sağlar.

Çoklu atama için en yaygın kullanılan yöntem sürekli değişkenler için, Markov Chain Monte Carlo (MCMC) algoritmasıdır ve ataması yapılacak verilerdeki değişkenliği dikkate alır (Cheung, 2007; Finch ve Margraf, 2008). Yöntem doğrusal regresyona dayanır ve çoklu normallik varsayımına dayalı beklenti maksimizasyonu algoritması ile benzerdir yani tekrarlı doğrusal regresyon ataması yapar. Aradaki temel fark, regresyonla değer atama sonucunda ortaya çıkan hata varyansı ile başa çıkmada kullandıkları yaklaşımdır. Beklenti maksimizasyonunda varyans ve kovaryans için değer atama yönteminden sonra düzeltme işlemi uygulanırken, çoklu atamada değer atama işlemi sırasında seçkisiz varyasyon (random variation) işlemi uygulanır. Bu işlemde, lineer regresyona dayalı tahmin değerleri üretildikten sonra her regresyon denklemi için hata dağılımı belirlenir. Belirlenen hatalar, ataması yapılacak verileri üretmek için her regresyon denklemine eklenir. Bu da varyans

tahmininde yanlılığı azaltır (Allison, 2003; Allison, 2009). Belirtilen formül şu şekildedir;

$$Y=a+bX+S_{x,y}E$$

E= Standart normal dağılımdan seçkisiz olarak belirlenen bir değer

$S_{x,y}$ = Regresyon modeli için belirlenen hata varyansı

Eşitlikten anlaşılacağı üzere veri atama işlemi için her seferinde farklı bir değer elde edilecektir. Bu formülden dolayı çoklu atama işleminde birden fazla değer ataması yapılmakta ve her atama işlemi sonucunda elde edilen parametre tahminleri birleştirilerek genel bir tahmin değeri oluşturulmaktadır (Dural, 2010).

7. Tam bilgi en çok olabilirlik- TBEO (full information maximum likelihood-Direct Maximum likelihood): Beklenti maksimizasyonuna benzer olup parametre tahmininde model temelli bir yaklaşımdır (Olinsky ve diğerleri, 2003). Model parametreleri ve standart hataların var olan veri üzerinden tek seferde hesaplandığı bir yöntemdir (Graham, 2009). AMOS, LISREL, EQX, MX, MPLUS gibi yapısal eşitlik modellemesinin yapıldığı programlarda kolayca uygulanabilmektedir (Allison, 2009; Enders, 2001 (a)). Yöntemde model parametreleri, kayıp verileri silme (Ginkel ve diğerleri, 2007) ve kayıp veri atama işlemi uygulanmadan tahmin edilir, yani bir atama veya silme yöntemi değildir. Gözlemlenen bütün bilgiler kullanılarak en çok olabilirlik yöntemi ile parametre tahmini yapılır (Acock, 2005; Enders, 2001(b)). Örneğin, bir veri setinde 5. durum kayıp veri içeriyor olsun. Yöntem, sadece 5. durum için çoklu normallığı varsayarak ve gözlemlenen değişkenleri kullanarak durum bazında olabilirlik fonksiyonunu hesaplar. TROK ve ROK mekanizmalarında diğer yöntemlere göre hem yansız hem de daha etkili sonuçlar vermektedir (Enders ve Bandalos, 2001).

Yöntemde kayıp veriler ve parametre tahminleri ile ilgili işlemler tek adımda gerçekleştirilir ve süreçte kovaryans matrisi yerine ham veriler kullanılır. Ham

veride yer alan bilgiler tek seferde okunarak, en çok olabilirlik yöntemiyle kayıp veri ile ilgili gerekli tahminler yapılır. (Graham, 2012). Çoklu normallik varsayımına dayanan modelde ortalama ve kovaryans matrisi, durum bazında özelleştirilen modeldeki parametrelerin bir fonksiyonu olarak ifade edilir. Olabilirlik fonksiyonu modeldeki parametrelere göre doğrudan maksimize edilir ve süreçte standart hata kestirimleri de iyi bir şekilde hesaplanmaktadır (Allison, 2009; Graham, 2012).

Bu yöntemle ilgili belirtilmesi gereken bazı noktalar vardır. Bunlardan biri yöntem, genel bir yöntem olup ortalama tahmini, kovaryans matrisi, çoklu regresyon ve yapısal eşitlik modeli gibi birçok analizde kullanılabilir. Bunun yanı sıra beklenti maksimizasyonu algoritmasından farklı olarak, standart hata kestirimleri bootstrap yapmaya gerek kalmadan analizlerle direkt elde edilir (Enders, 2001(b)). Liste bazında ve çiftler bazında silme yöntemlerinden farkı ise sadece tam durumları içeren verileri kullanmaktansa, tüm durumları kullanmasıdır (Ginkel ve diğerleri, 2010).

Yöntemin dezavantajları ise şu şekildedir; yöntem, çoklu atama yöntemlerinde olduğu gibi biraz karışıktır; hem karışık bir istatistiksel hesaplama sürecine sahiptir hem de standart olmayan istatistiksel programlar gerektirir (Ginkel ve diğerleri, 2010). Yöntem yapısal eşitlik modelinde kullanıldığında ki-kare testi normal bir şekilde hesaplanamamaktadır. Çünkü tüm örneklem için uygulanacak tek bir N değeri yoktur. Bir diğeri, TBEO yöntemi altında birçok fit indeksi hesaplanmasına rağmen GFI gibi tahmin süreci için gerekli olan bazı ortalama değerleri tanımsız olmaktadır. Yöntemle ilgili diğeri bir sorun ise çiftler bazında silme yönteminde olduğu gibi kovaryans matrisleri tanımsız olabilmektedir (Enders, 2001(b)). Bunların yanı sıra, yöntem temel bileşenler analizi ve varyans analizi (ANOVA) gibi yöntemlerde kullanılamamaktadır (Ginkel ve diğerleri, 2010).

8. Hot Deck ataması- HD (Hot Deck Imputation): Veri matrisindeki kayıp verilerin benzer gözlemlerle doldurulduğu tekniktir (Brown ve Kros, 2003; Kaspar, 2011; Oğuzlar, 2011). Benzer gözlemlerden kasıt, analizi yapan kişi

tarafından belirlenen bir grup benzer deęişken arasından seçilen gözlemlerdir (Buhi ve dięerleri, 2008). Yöntem, uzun yıllardır kullanılan bir yöntem olup LBS, ÇBS ve OA yöntemlerinden daha üstündür (Oğuzlar, 2001).

Hot deck atamasının uygulanması için 2 yöntem vardır. Yöntemlerden birincisinde kayıp veri için birçok durum havuzu vardır ve buna donör havuzu denir. Donör havuzunda bulunan durumlar ile kayıp veri içeren durumlar arasında çeşitli deęişkenler açısından benzerlik vardır. Donör havuzundan rastgele bir durum seçilerek kayıp veri ile yer deęiştirilir. Bu yöntem rastgele hot deck ataması (random hot deck imputation) olarak adlandırılır. Hot deck ataması için dięer bir yöntem de en yakın donör ile deęişim yapmaktır. Bu yöntem de belirleyici hot deck ataması olarak adlandırılır. Yöntemin iyi bir şekilde çalışabilmesi için büyük bir örneklem gereklidir. Uygun donör seçimi için tüm olası deęişkenler açısından alıcı (kayıp veri içeren deęişkenler), benzer durumlarla eşleştirilir. Bu atama sırasında göz önünde bulundurulması gereken iki kriter vardır; dışarıdan seçilen deęişken ile ataması yapılacak deęişkenin benzer olması ve seçilen deęişkenin kayıp veri içeren veya içermeyen ikili deęişkenle ilişkili olup olmamasıdır (Andridge ve Little, 2010; Cheema,2012).

Hot deck ataması deęişkenlerin ölçüm düzeylerinin korunması (kategorik ve sürekli olarak), kavramsal basitlięi ve atıf sonucunda tamamlanmış veri matrisi elde edilmesi açısından avantajlı bir yöntemdir. Yöntemin en önemli dezavantajı “benzerlik” kavramının tanımlanmasındaki güçlükten dolayı kayıp veriler için standart bir yol sağlayamamasıdır. Benzerlięin belirlenebilmesi için donör durumların seçimini yapabilecek bir yazılım gerekmektedir (Oğuzlar, 2001).

9. Cold Deck ataması- CD (Cold deck imputation): Hot deck atamasında olduęu gibi bir kaynaktan kayıp veri yerine atama yapılır. Burada kayıp veri ataması için kullanılan kaynaklar önceki çalışmalardan veya uygulamalardan elde edilen deęerlerdir. Bu kaynaklardan elde edilen deęerler kayıp veriler yerine

atanır (Brown ve Kros, 2003; Lakshminarayan ve diğeri, 1999). Temel prensip kayıp verileri doldurmak için güncel verilerden faydalanmaktır (Baygöl, 2007). Ancak, cold deck ataması çok tercih edilen bir yöntem olmamakla birlikte genellikle hot deck ataması yapılabilmesi için başlangıç verilerini sağlamada kullanılmaktadır (Brown ve Kros, 2003).

Gözlemlenen değişkenlerden gizil özelliklerin ölçülmesi amacıyla ölçme araçlarının geliştirilmesi, uyarlanması veya var olan araçların güncellenmesi söz konusu olmaktadır. Bu süreçte kullanılan ölçme araçlarının geçerliği ve güvenilirliğinin sağlanması gerekir. Geçerlik ve güvenilirliği test etmek amacıyla geliştirilen istatistiksel programlar tam veri setlerine göre geliştirildiği için verilerde kayıp olması geçerlik ve güvenilirliği etkilemektedir (Akbaş, 2014). Benzer şekilde betimsel istatistik sonuçları da bu durumdan etkilenmektedir (Bal, 2003; Demir, 2013). Çalışmada betimsel istatistik, güvenilirlik ve geçerlik değerleri açısından yöntem karşılaştırması amaçlandığı için bu konu ile ilgili açıklamalara aşağıda yer verilmiştir.

#### 2.4. Betimsel İstatistikler

Bir grup objenin belirli bir özelliği üzerinde çalışılırken, bu özellik ile ilgili veriler toplanır. Veri kümesi bazen az bazen de çok fazla eleman içerebilir. Veri kümesinin elemanlarının az olduğu durumda tüm veri üzerinden çalışılırken, çok fazla olduğu durumda bu elemanları temsil edecek bir grup üzerinden çalışılır. Tüm elemanların oluşturduğu kümeye evren adı verilirken, bu evreni temsil eden gruba da örneklem adı verilir. Çoğunlukla araştırmalarda amaç, evrenin özelliklerini kestirmektir. Ancak bazı durumlarda evrene ulaşmak zor olduğu için örneklemden yararlanılarak evren konusunda çıkarımlarda bulunulur (Baykul, 2010).

Araştırmalarda toplanan veriler ilk halleri ile hamdırlar. Örneğin, bir öğrencinin aldığı ham puan doğrultusunda, öğrencinin durumu veya gruptaki konumu hakkında yorum yapmak zordur. Bundan dolayı ham puanların anlamlı hale getirilerek yorumlanması gerekir (Tan, 2012). Ham puanların anlamlı hale getirilmesi için tablo ve

grafik şeklinde düzenlenmesi ve ortalama, standart sapma, basıklık, çarpıklık vb. ölçütlerin hesaplanması gerekir (Baykul, 2010).

Araştırmalarda evren hakkında çıkarımlarda bulunmak için belirtilen merkezi eğilim, merkezi dağılım vb. ölçütlerin hesaplanması gerekir. Araştırmalarda genellikle merkezi eğilim ölçülerinden ortalama; değişim ölçülerinden standart sapma ve puan dağılımının normalliğini kontrol etmek amacıyla basıklık ve çarpıklık değerleri incelenmektedir. Bundan dolayı bu araştırma için betimsel istatistik değerlerinden ortalama, standart sapma, basıklık ve çarpıklık değerleri, kayıp veri çalışması için ele alınmıştır.

## 2.5. Güvenirlik ve Geçerlik

Ölçme, “nesne ya da bireylerin ilgilenilen özelliklerinin derecesini sayı, simge ya da sıfatlarla gösterme/ eşleştirme işlemi ve süreci” olarak tanımlanmaktadır (Erkuş, 2006). Ölçmenin amacı yapılan deney ve gözlemlerden geçerli ve güvenilir veri sağlamaktır (Baykul,2010). Geçerlik ve güvenirlilik ölçme aracında bulunması gereken temel niteliklerdendir. Güvenirlilik, ölçme aracının, ölçülmek istenen özelliği tutarlı olarak ölçmesidir. Geçerlik ise ölçme aracının ölçülecek özelliği başka özelliklerle karıştırmadan tam ve doğru olarak ölçmesidir yani ölçme aracının amacına hizmet etme derecesidir (Tekin,2012). Aslında geçerliğe tek bir tanım yapmak doğru olmaz. Bunun yerine testin kullanım amacına uygun kanıtlar ortaya koymak demek daha doğru olacaktır (Baykul, 2010).

Bu çalışma kapsamında karşılaştırma yapmak amacıyla güvenirlilik çalışması için Cronbach Alpha güvenirliliği kullanılmıştır. Cronbach alpha katsayısı ölçek puanlamasının çoklu olduğu durumda hesaplanan bir güvenirlilik değeridir (Ercan ve Kan, 2004). Çalışma kapsamında kullanılan ölçek 1-4 şeklinde çoklu puanlanan bir ölçek olduğu için güvenirlilik hesabı için Cronbach Alpha katsayısının kullanılması uygun bulunmuştur. Çalışma kapsamında ele alınan bir diğer konu da ölçülen özelliğin yapısını test etmede kullanılan açıklayıcı faktör analizidir. Faktör analizinin amacı,

maddelerin gerçekte ölçülmek istenen yapıyı ölçüp ölçmediğini belirlemek ve belirlenen yapıya ilişkin bağımsız faktörleri ortaya çıkarmaktır (Büyüköztürk, 2002).

Ölçme aracının ölçmeyi hedeflediği durumu veya olguyu ölçmesi ve tekrarlayan ölçümlerde benzer sonuçların elde edilmesi, ölçme aracının genellenebilirliği ile alakalıdır. Bilimsel araştırmalarda da somut yargı ve kararlara ulaşabilmek için güvenilirlik ve geçerliği kanıtlanmış yani genellenebilir ölçme araçlarından elde edilen verilerin değerlendirilmesi gerekir (Çakmur, 2012). Kayıp verilerin geçerlik ve güvenilirliği, doğal olarak çalışmanın genellenebilirliğini etkilediği düşünülürse (McKnight ve diğerleri, 2007), kayıp veri yöntemlerinin geçerlik ve güvenilirliğe etkisinin incelenmesi önem kazanmaktadır.

## 2.6. İlgili Araştırmalar

Kayıp verilerle ilgili yapılan çalışmalar 80 yıllık uzun bir sürece dayanmaktadır. Bu süreçte yapılan çalışmalar doğrultusunda, çoğu araştırmacı bu yöntemler arasında çoklu atama ve en çok olabilirlik temelli yöntemleri önermektedir. Bu yöntemler “modern” yaklaşımlar olarak adlandırılmakla birlikte daha güçlü teorik altyapı sağlaması, daha az sınırlayıcı varsayımlara sahip olması, yanlılığı azaltması ve daha büyük istatistiksel güç sağlaması açısından diğer yöntemlere göre daha avantajlıdır (Enders, 2013).

Literatür incelendiğinde kayıp veri ile ilgili birçok çalışmaya rastlanmaktadır. Bunlardan biri Allison (2003) tarafından kayıp verinin yapısal eşitlik modeli üzerindeki etkisi ile ilgili yapılan bir çalışmadır. Bu çalışmada, silme ve basit atama gibi geleneksel yöntemlerin standart hata ve/veya parametre tahminlerinde yanlılığa neden olduğu, en çok olabilirlik ve çoklu atama gibi yöntemlerin parametre tahminlerinde daha tutarlı sonuçlar verdiğini belirtmiştir. Benzer sonuçlar Allison (2009)’un çalışmasıyla da desteklenmiştir.

Alosh (2009) yapay verileri kullandığı çalışmasında 30 ve 60 kişilik yapay veri seti üreterek %10 ve %20 oranlarında kayıp durumları oluşturmuştur. Çalışmada TROK, ROK, IE ve ROK mekanizması altında son gözlemi ileri taşıma yöntemini kullanarak kayıp veri mekanizmaları karşılaştırılmıştır. Çalışma sonucuna göre en çok yanlılığın IE mekanizması altında üretildiğini belirtmiştir. Benzer şekilde Buhi ve diğerleri (2008) kayıp veri mekanizmaları ve kayıp veri yöntemleri ile ilgili yaptıkları çalışmada en fazla yanlılığın IE mekanizması ile elde edildiği; TROK mekanizmasından ise uygulanan kayıp veri yöntemleri içinde yansız parametre tahminleri elde edildiği belirtilmiştir. MAR mekanizması altında ise en çok olabilirlik ve çoklu atama gibi modern yöntemlerin neredeyse tam veriye yakın sonuçlar verdiği belirtilmiştir.

Akbaş (2014) farklı örneklem büyüklüğü ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerini kayıp veri yöntemleri ile incelediği çalışmasında LBS yönteminin çeşitli sorunlara yol açabileceğini belirtmiştir. ÇA ve BM yöntemlerinin ise yüksek performans gösterdiğini ancak tüm durumlar için önerilebilecek tek bir yöntemin olmadığını belirtmiştir.

Pigott (2001) ise sağlık alanında yaptığı çalışmasında araştırmacıların genellikle liste bazında silme, çiftler bazında silme veya tekli atama yöntemlerini tercih ettiklerini belirtmiştir. Ancak beklenti maksimizasyonu ve çoklu atama gibi yöntemlerin uygulamada bazı zorlukları olmasına rağmen daha iyi sonuçlar verdiğini belirtmiştir. Larsen (2011) ise çoklu atama ve tam bilgi en çok olabilirlik yöntemlerini TROK ve ROK varsayımları altında karşılaştırdığı simülatif çalışmasında standart hata tahminlerini daha doğru yapmasından dolayı tam bilgi en çok olabilirlik yönteminin çoklu atama yöntemine göre daha doğru sonuçlar verdiğini belirtmiştir.

Bal (2003) kayıp veri yöntemlerini, sağlık alanında yaptığı bir çalışmada karşılaştırmıştır. Yapay olarak üretilen veri setlerinde LBS, ÇBS, RA ve BM yöntemlerini karşılaştırmış ve tam veri setine en yakın sonuçların BM yönteminden elde edildiğini belirtmiştir. Sağlık alanında yapılan bir başka çalışmaya göre LBS, ÇBS, RA, HD, CD, BM, ÇA ve son gözlemi ileri taşıma yöntemleri simülatif verilerle karşılaştırılmış ve yöntemler arasında önemli bir farklılığın bulunmadığı; silme ve

atama yöntemlerinin veri setine uygun kullanıldığı sürece anlamlı sonuçlar verdiği belirtilmiştir (Baygöl, 2007).

Baraldi ve Enders (2010)'ın "modern" kayıp veri yöntemlerini karşılaştırdığı çalışmasına göre de LBS, OA gibi geleneksel yöntemlerdense EÇO ve ÇA yöntemlerinin daha iyi sonuçlar verdiği belirtilmiştir. Geleneksel yöntemlerin, yanlış parametre tahminlerine neden olmasından dolayı modern yöntemlerin kullanılması önerilmiştir. Cool (2000) de LBS, ÇBS, OA ve RA yöntemlerini karşılaştırdığı çalışmasında silme yöntemlerinin örneklem büyüklüğünü düşürdüğü ve buna bağlı olarak populasyon parametrelerinin tahmininin kesin değerlerden uzak yapıldığını belirtmiştir. OA ve RA gibi atama yöntemlerinin de örneklem büyüklüğünü ve buna bağlı olarak istatistiksel gücü korumasından dolayı silme yöntemlerine göre avantajlı olduğu belirtilmiştir.

Chen, Wang ve Chen (2012) kayıp veri yöntemlerini açıklayıcı ve doğrulayıcı faktör analizleri üzerindeki etkisine göre karşılaştırmıştır. LBS, SO, NDE, DDK, BM ve RA yöntemlerini kullandığı çalışmada BM yöntemi en iyi sonucu vermiştir. Kayıp veri oranının %20'den düşük olduğu durumda atama yöntemleri arasında istatistiksel olarak anlamlı fark gözlemlenmezken, kayıp veri oranının %30'dan büyük olduğu durumlarda seri ortalaması ve noktanın doğrusal eğimi yöntemlerinin kullanılması önerilmiştir.

Cheung (2007) çalışmasında LBS, OA, BM, ÇA ve TBEO yöntemlerini TROK varsayımı altında yapay verileri kullandığı bir çalışmayla karşılaştırmıştır. TBEO ve LBS yöntemleri en iyi çalışan yöntemler olurken; BM ve ÇA yöntemlerinde standart hataların olduğundan düşük hesaplandığı bulunmuştur. OA yönteminin ise kayıp veri oranının düşük olduğu durumlarda iyi çalıştığı belirtilmiştir.

Demir (2013) çoktan seçmeli testlerde madde ve test parametreleri üzerinde farklı kayıp veri yöntemlerini karşılaştırmıştır. Kayıp verilerin ihmal edilebilir olmadığı durumlarda karşılaştırılan yöntemler arasında silme ve 0 atama (iki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde kullanılan kayıp veri yöntemi) yöntemlerinin uygun yöntemler olmadığı; basit atama yöntemlerinde yanlış parametre

tahminleri yapıldığı; en çok olabilirlik ve çoklu veri atama yöntemlerinin ise kayıp verinin ihmal edilebilir olmadığı durumlarda en uygun yöntemler olduğu belirtilmiştir. Köse ve Öztemur (2014)'ün kayıp verilerin t-testi ve ANOVA parametreleri üzerine etkisinin araştırıldığı çalışmasında silme, ortalama atama, regresyon ataması ve beklenti maksimizasyonu yöntemleri karşılaştırılmıştır. Düşük birimli veri setlerinde regresyon ataması ve beklenti maksimizasyonu yöntemlerinin iyi sonuçlar verdiği; yüksek birimli veri setlerinde ise regresyon ataması ve yerine ortalama koyma yöntemlerinin tam veri setiyle daha tutarlı sonuçlar verdiği belirtilmiştir.

Jing (2012)'nin beklenti maksimizasyonu, regresyon ataması, skotastik regresyon ataması ve çoklu atama yöntemlerini karşılaştırdığı çalışmasında, çok fazla kayıp veri varsa çoklu atama yönteminin kullanılması önerilmiştir. Bu bulgu Kang (2013)'ün çalışmasıyla da desteklenmiştir. Kaspar (2011) de gerçek veri setleri ile sıklıkla kullanılan ortalama atama, medyan atama, beklenti maksimizasyonu, regresyon ataması, hot-deck atama yöntemleri ile propensity skor yöntemlerini karşılaştırmıştır. Yöntemlerin etkinliğine tam veri seti ile karşılaştırarak bakılmıştır. Propensity skor yöntemi kayıp veri miktarının düşük olduğu durumlarda diğer yöntemlere göre daha iyi sonuç verirken, tüm kayıp veri durumlarında beklenti maksimizasyonu ve propensity skor yöntemlerinin daha iyi sonuçlar verdiği belirtilmiştir.

Literatürde eğitim, sağlık, psikoloji, sosyoloji vb. alanlarda bu çalışmalara benzer kayıp veri yöntemleri, kayıp verinin ortaya çıkardığı problemler, bu problemlerin nasıl değerlendirileceği vb. konularda çalışmalar mevcuttur. Ancak, bu yöntemleri çalışmalara uygulamaya geldiğinde araştırmacılar yetersiz kalmaktadır. Bunun temel sebeplerinden biri, araştırmacıların kayıp verinin mantığını anlamak için yetersiz bir matematik ve istatistik bilgisine sahip olmasıdır. Diğer bir neden kayıp veri yöntemleri için kullanıcı uyumlu araçların sınırlı sayıda olmasıdır. Aslında bu konuyla ilgili birçok istatistik programı olmasına rağmen, yöntemlerin uygulanması ile ilgili yeterli sayıda doküman yoktur. Başka bir sebep ise kayıp veri ile ilgili literatürün, bu konuyla ilgili uygulamalar konusunda yeterince etkisinin olmamasından dolayı araştırmacıların bu yöntemleri çalışmalarında kullanma zorunluluğu hissetmemesinden kaynaklanmaktadır (McKnight ve diğerleri, 2007).

## BÖLÜM III

### 3. Yöntem

Bu bölümde araştırmanın modeli, veri toplama aracı, çalışma grupları, verilerin düzenlenmesi, verilerin analizi ve araştırmada kapsamında ele alınan kayıp veri yöntemlerine yer verilmiştir.

#### 3.1. Araştırma Modeli

Bu araştırma, farklı koşullar altında çeşitli kayıp veri yöntemleri kullanılarak, bu yöntemlerin betimsel istatistik, güvenilirlik ve geçerlik üzerindeki etkilerinin karşılaştırmalı olarak incelendiği temel araştırma niteliğinde bir çalışmadır. Temel araştırmalar var olan bilgiye yenilerini eklemek amacıyla gerçekleştirilen teorik veya deneysel nitelikte çalışmalardır (Karasar, 2007). Bu doğrultuda, farklı kayıp veri yöntemlerinin kullanılması ile betimsel istatistikler, güvenilirlik ve geçerlik bağlamında elde edilen bulgular örneklem büyüklüğü ve kayıp veri oranı değişkenlerinin farklı düzeyleri üzerinden, tek boyutluluk ve normal dağılım koşulu altında açılımcı bir yaklaşımla karşılaştırılarak incelenmiştir.

#### 3.2. Veri Toplama Aracı ve Veriler

Araştırmada Uluslararası Öğrenci Değerlendirme Programı- PISA (Programme for International Student Assessment) verilerinden yararlanılmıştır. PISA verileri temel olarak başarı testleri, öğrenci ve okul anketlerinden oluşmaktadır. Ölçülen beceriler fen, matematik ve okuma alanlarında temel, orta ve ileri düzeyde değerlendirilmektedir

(Şirin ve Vatanartıran, 2014). Araştırma için PISA verilerinin seçilmesinin nedeni örneklem oluşturma ve veri toplama işlemleri esnasında uyulması zorunlu belirli mekanizmaların olmasından dolayı elde edilen bulguların yüksek derecede geçerlik ve güvenilirliğe sahip olmasıdır (Anıl, 2011). Bu araştırmada da, betimsel istatistiklerin yanı sıra geçerlik ve güvenilirlik açısından kayıp veri yöntemleri arasında karşılaştırma yapılacağı için PISA verileri tercih edilmiştir.

Araştırma kapsamında PISA 2012 öğrenci ölçeklerinden “Matematik Çalışma Etiği (Math Work Ethic)” ölçeği seçilmiştir. Araştırmada bu ölçeğin seçilmesinin nedeni; 3 yılda bir tekrarlanan PISA sınavlarında matematik, okuma becerisi ve fen alanlarından biri uygulamanın ağırlıklı değerlendirme alanını oluşturmaktadır ve 2012 yılında da ağırlıklandırılan alan matematik olmuştur (Milli Eğitim Bakanlığı [MEB], 2011). Bundan dolayı matematik ölçeklerinden tek faktörlü olan bir ölçek rastgele seçilmiştir. Ölçek, “Kesinlikle katılıyorum (1)”, “Katılıyorum (2)”, “Katılmıyorum (3)”, “Kesinlikle katılmıyorum (4)” şeklinde 4’lü likert tipinde ve 9 maddeden oluşmaktadır.

Araştırma için PISA veri tabanından PISA 2012 veri dosyaları alınmıştır ve araştırma kapsamı için Türkiye örneklemini seçilmiştir. Kod kitaplarından yararlanılarak veriler yeniden düzenlenmiştir. Buna göre, verilerden “7 (N/A)- Hiç bir sorunun uygulanmadığı (Baskı veya çeviri hatasından dolayı cevaplanmayan sorular)”, “8 (Invalid)- Geçersiz cevap (birden fazla seçeneğin işaretlenmesi)” ve “9 (Missing)- Kayıp” veriler çıkartılarak tam veri veri dosyasına son hali verilmiştir.

### 3.3. Çalışma Grupları ve Verilerin Düzenlenmesi

PISA 2012 sınavına Türkiye’den 4848 öğrenci katılmıştır. Araştırma için gerekli olan veri setlerini oluşturmak amacıyla öncelikle 4848 veri arasından “Matematik Çalışma Etiği” ölçeğine göre kayıp, geçersiz cevap bulduran ve hiçbir soruyu cevaplamayan bireyler çıkartılarak 3127 kişilik tam veri grubu elde edilmiştir. Daha sonra 3127 kişi arasından, araştırmanın amacı doğrultusunda, SPSS 20.0 paket programı kullanılarak rastgele 200, 500 ve 1000 kişilik, tam veri setine sahip, çalışma grupları

seçilerek belirli oranlarda kayıp veri durumları oluşturulmuştur. Belirtilen işlemle ilgili sonuçlar tablo 3.1’de belirtilmiştir.

**Tablo 3.1.** Yüzdelerle göre kayıp veri miktarları

Kayıp veri oranı	%0	%5	%10	%20
N	200	190	180	160
	500	475	450	400
	1000	950	900	800

Tablo 3.1.’de görüldüğü üzere n=200, n=500 ve n=1000 kişilik 3 farklı çalışma grubu vardır. Tam veriye sahip 3 farklı çalışma grubunun her birinden rastgele ve farklı oranlarda (%5, %10, %20) veri silinmiştir. 200 kişilik tam veri grubundan %5, %10 ve %20 oranlarında veri silindiğinde sırasıyla 190, 180 ve 160 kişilik veri grubu elde edilmiştir. 500 kişilik tam veri grubundan %5, %10 ve %20 oranlarında veri silindiğinde sırasıyla 475, 450 ve 400 kişilik veri grubu elde edilmiştir. 1000 kişilik tam veri grubundan %5, %10 ve %20 oranlarında veri silindiğinde ise sırasıyla 950, 900 ve 800 kişilik veri grubu elde edilmiştir. Sonuç olarak toplamda 12 adet veri seti elde edilmiştir.

#### 3.4. Verilerin Analizi

Verilerin analizi sürecinde, öncelikle elde edilen 12 adet veri setinin her biri için betimsel istatistik değer kestirimleri yapılmış ve dağılımların normalliği test edilmiştir. Farklı sayıda kişilerden oluşan ve farklı oranlarda kayıp veri içeren veri setlerinden toplam puanlar için elde edilen betimsel istatistik değerleri Tablo 3.2.’de verilmiştir.

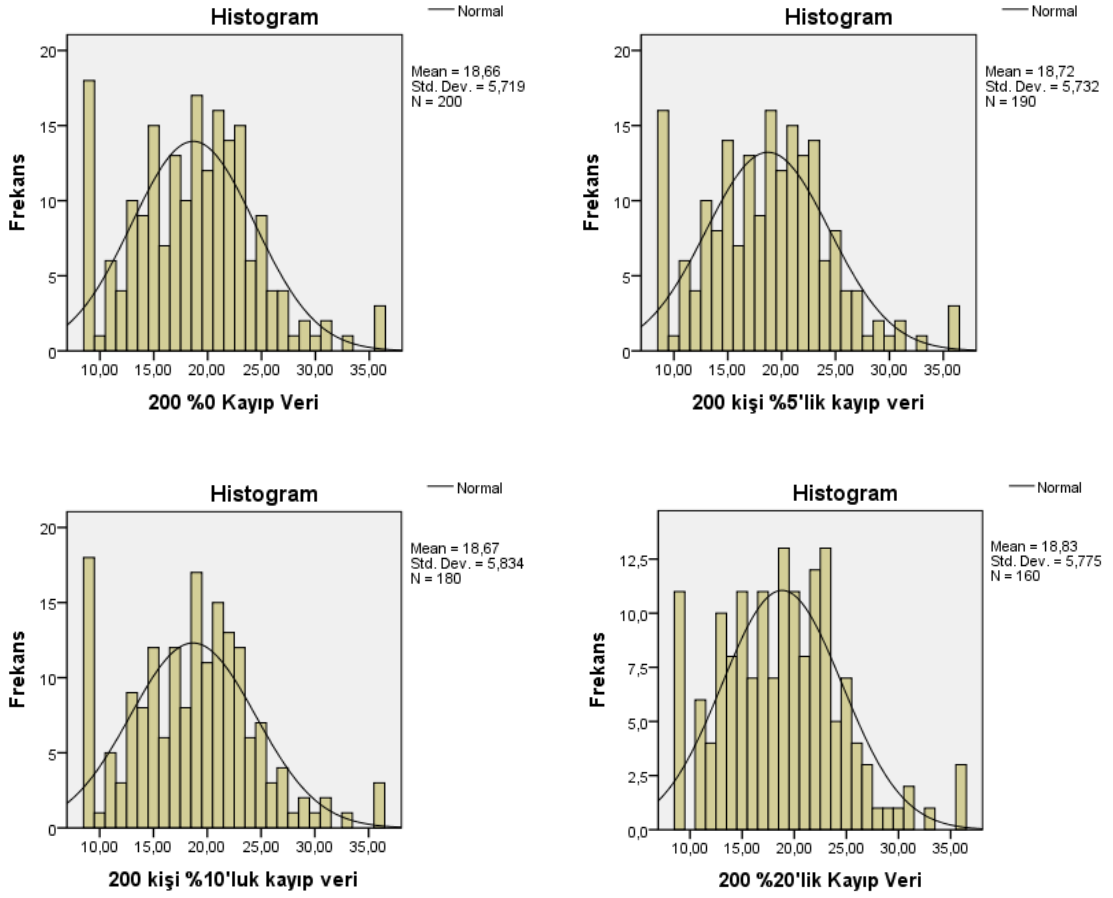
**Tablo 3.2.** Farklı sayıda kişi ve farklı oranlardaki kayıp veri içeren veri setlerinden toplam puanlar için elde edilen betimsel istatistikler

Kayıp veri oranı	n	Ortalama ( $\bar{X}$ )	Standart sapma (S)	Min.	Maks.	Çarpıklık	Basıklık
%0	200	18,660	5,719	9,00	36,00	,313	,214
%5	190	18,721	5,732	9,00	36,00	,241	,339
%10	180	18,666	5,833	9,00	36,00	,324	,232
%20	160	18,831	5,774	9,00	36,00	,442	,313
%0	500	18,724	5,600	9,00	36,00	,571	,590
%5	475	18,768	5,643	9,00	36,00	,586	,587
%10	450	18,646	5,573	9,00	36,00	,572	,556
%20	400	18,735	5,643	9,00	36,00	,634	,728
%0	1000	18,579	5,732	9,00	36,00	,458	,321
%5	950	18,608	5,756	9,00	36,00	,458	,340
%10	900	18,687	5,758	9,00	36,00	,479	,375
%20	800	18,435	5,836	9,00	36,00	,494	,253

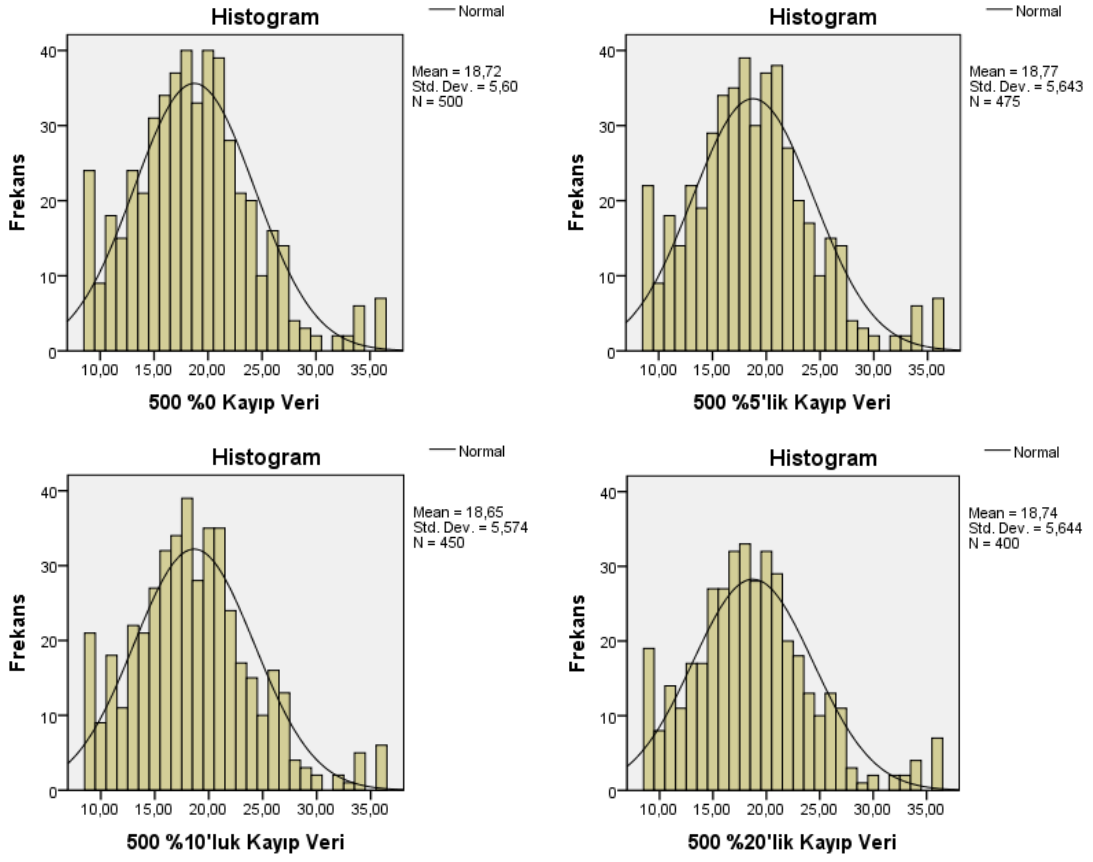
Farklı sayıda veri setleri ve farklı kayıp veri oranlarına göre elde edilen toplam puan ortalamaları 18,435 ile 18,831 değerleri arasında değişirken, standart sapma değerleri 5,573 ile 5,836 arasında değişen değerler almıştır. Farklı oranlardaki kayıp veriler için minimum değer 9,00 maksimum değer ise 36,00'dır.

Toplam puanlara göre çarpıklık değerleri ise ,241 ile ,634 arasında, basıklık değerleri ,214 ile ,728 arasında değerler almıştır. Çarpıklık katsayısının -1 ve +1 arasında değerler aldığı durumda, puanlar normalden önemli bir sapma göstermemektedir (Büyüköztürk, 2012; Huck, 2012). Benzer şekilde basıklık değerinin de -1 ile +1 arasında değişen değerler alması normal dağılımın göstergelerinden biridir (Huck, 2012). Tablo 3.2'de görülen toplam puanlara göre basıklık ve çarpıklık değerleri -1 ile +1 arasında değişen değerler aldığı için verilerin normal dağıldığı söylenebilir.

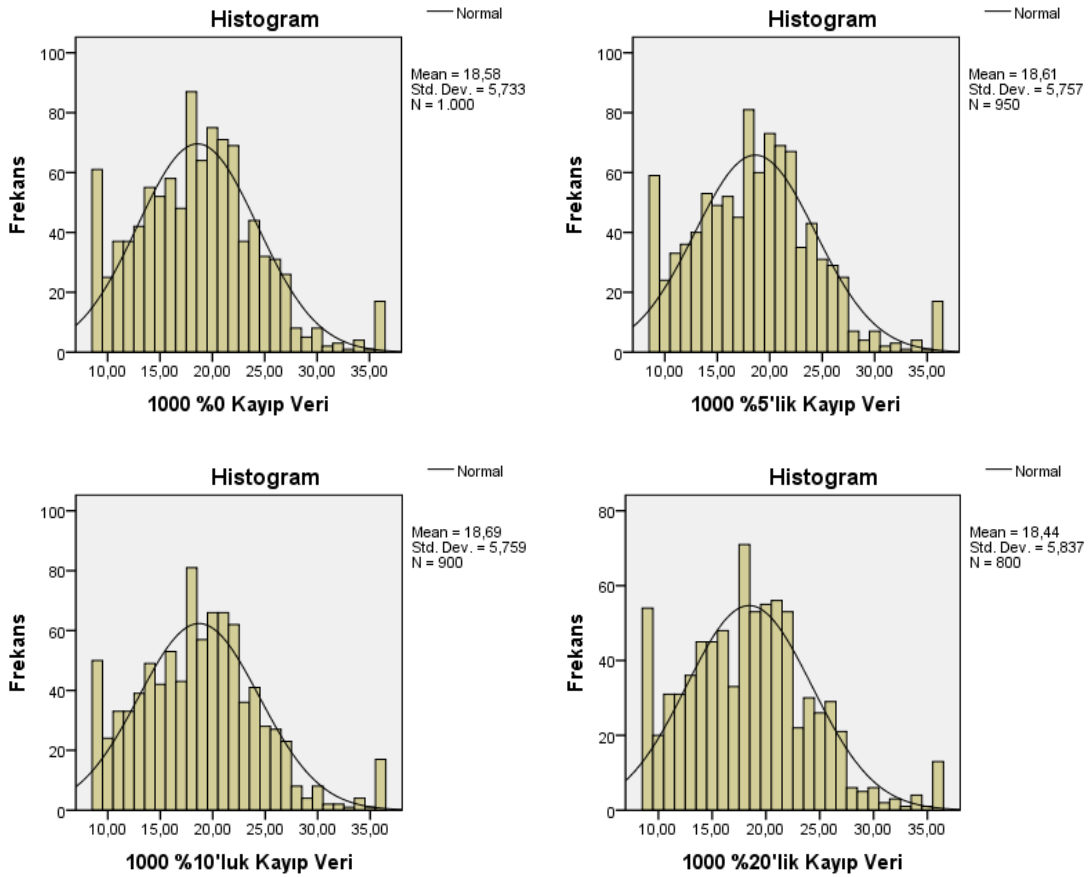
Verilerin dağılımını gösteren histogram grafikleri de Şekil 3.1, Şekil 3.2 ve Şekil 3.3'te de verilmiştir.



Şekil 3.1. Farklı oranlarda kayıp veri içeren 200 kişilik veri seti için histogram grafikleri



Şekil 3.2. Farklı oranlarda kayıp veri içeren 500 kişilik veri seti için histogram grafikleri



**Şekil 3.3.** Farklı oranlarda kayıp veri içeren 1000 kişilik veri seti için histogram grafikleri

Verilerin analizi sürecinin ikinci aşamasında 200, 500 ve 1000 kişilik çalışma grupları oluşturulduktan sonra, her bir çalışma grubundan %5, %10 ve %20 oranlarında veri silinerek yeni veri grupları oluşturulmuştur. Verilerin silinmesinde rastgelelik dikkate alınmış olsa da rastgeleliğin sağlandığını görmek amacıyla verilere Little'ın TROK testi uygulanmıştır. Bunun için SPSS 20.0 programında bulunan BM algoritmasına bağlı olarak hesaplanan Little'ın TROK testi kullanılmıştır. Verilere BM algoritması uygulanırken 2 varsayımın sağlanması gerekmektedir; bunlardan birincisi verilerin kesikli olmaması, ikincisi de verilerin normal dağılmasıdır (Oğuzlar, 2001). Araştırma verileri bu iki varsayımı da sağladığı için BM algoritması uygulanmış ve sonuçlar Tablo 3.3.'te verilmiştir.

**Tablo 3.3.** Little'ın TROK testi sonuçları

N	Kayıp veri oranı	<i>p</i>
200	%5	,99
	%10	,74
	%20	,88
500	%5	,37
	%10	,54
	%20	,36
1000	%5	,80
	%10	,23
	%20	,80

Tablo 3.3.'teki Little'ın TROK testi sonuçları incelendiğinde *p* değerlerinin ,23 ile ,99 arasında değerler aldığı görülmektedir. Bu değer 0,05'ten büyük olması verilerin TROK mekanizmasına uyduğunun göstergesidir (IBM,2012). Araştırmada kullanılacak çalışma grupları için bu değer 0,05'in üzerinde olduğu için verilerin TROK mekanizmasına uyduğu görülmektedir.

Verilerin betimsel istatistik ve TROK sonuçlarının incelenmesi neticesinde verilerin ilgili analizler için uygun olduğu sonucuna varılmıştır.

Analizlerde 200, 500 ve 1000 kişilik çalışma grupları için oluşturulan tam veri setlerinden elde edilen betimsel istatistikler, güvenilirlik ve geçerlik kanıtları karşılaştırmalarda referans değerler olarak ele alınmıştır. Farklı kayıp veri yöntemleri neticesinde elde edilen betimsel istatistikler, güvenilirlik ve geçerlik kanıtları ile referans değerler karşılaştırılmıştır.

Araştırmanın birinci alt probleminde, farklı kayıp veri yöntemlerinden elde edilen betimsel istatistik değerlerinin referans gruptan elde edilenlerden farklı sonuçlar verip vermediği öncelikle betimsel düzeyde karşılaştırılmış, hesaplanan ortalama değerleri arasında manidar farklılığın bulunup bulunmadığı ise ilişkili örneklem için t-testi ile test edilmiştir.

Araştırmanın ikinci alt probleminde, kayıp veri yöntemlerinden elde edilen sonuçlar tam veri setleri ile güvenilirlik açısından karşılaştırılmıştır. 200, 500 ve 1000 kişilik tam veri setleri için güvenilirlik katsayısı sonuçları tablo 3.4'te belirtilmiştir.

**Tablo 3.4.** Tam veri setleri için Cronbach Alpha katsayıları

Tam Veri Setleri	Cronbach Alpha Katsayısı
200	,910
500	,902
1000	,910

Tablo 3.4'te cronbach alpha sonuçlarına bakıldığında 200 kişilik veri seti için ,910; 500 kişilik veri seti için ,902; 1000 kişilik veri seti için ise ,910 değeri elde edildiği görülmektedir.

Araştırmanın ikinci alt probleminde, farklı kayıp veri yöntemlerinden elde edilen güvenilirlik katsayılarının referans gruptan elde edilenden farklı sonuçlar verip vermediği öncelikle betimsel düzeyde karşılaştırılmış, manidar farklılığın bulunup bulunmadığı ise Fisher'in z testi ile test edilmiştir.

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

Formülde yer alan  $z_1$  ve  $z_2$ , güvenilirlik katsayısının z değerine karşılık gelen değerini,  $n_1$  ve  $n_2$  ise örneklem büyüklüğünü temsil etmektedir. Eşitlikten bulunan z değeri 0,05 anlamlılık düzeyinde -1,96 ile +1,96 arasında değerler alıyorsa bu durum değerler arasında manidar bir farklılığın olmadığı göstergesi sayılmaktadır (Akhun, 1994; Kenny, 1987).

Araştırmanın üçüncü alt probleminde, kayıp veri yöntemlerinden elde edilen sonuçlar tam veri setleri ile açımlayıcı faktör analizi sonuçları açısından karşılaştırılmıştır. 200, 500 ve 1000 kişilik tam veri setleri için açımlayıcı faktör analizi sonucu elde edilen madde faktör yük değerleri, açıklanan varyans ve özdeğerlerle ilgili sonuçlar tablo 3.5'te belirtilmiştir.

**Tablo 3.5.** Tam veri setleri için açımlayıcı faktör analizi sonuçları

Maddeler	200	500	1000
1	,71	,75	,75
2	,78	,79	,80
3	,77	,74	,77
4	,73	,72	,76
5	,80	,78	,77
6	,81	,77	,80
7	,77	,72	,77
8	,75	,74	,72
9	,75	,74	,72
<b>Açıklanan Varyans</b>	58,41	56,17	58,20
<b>Özdeğer</b>	5,26	5,06	5,24

Tam veri setleri için 9 maddelik ölçeğin açımlayıcı faktör analizi sonuçlarına göre tüm örneklem büyüklükleri için ölçeğin tek faktörlü yapısı korunmuştur. Tablo 3.5'te madde faktör yük değerleri incelendiğinde 200 kişilik veri seti için madde faktör yük değerlerinin ,71 ile ,81 arasında değiştiği görülmektedir. 500 kişilik veri seti için madde faktör yükleri ,72 ile ,79 arasında değişirken, 1000 kişilik veri seti için ,72 ile ,80 arasında değiştiği görülmektedir. Faktör yük değerinin ,45 ve üzerinde olması madde seçimi için iyi bir ölçüt olarak kabul edilmektedir (Büyüköztürk, 2012). Bundan dolayı 200, 500 ve 1000 kişilik tam veri setleri için madde faktör yükleri analizler için yeterli düzeydedir.

Açıklanan varyans değerleri incelendiğinde 200 kişilik veri seti için 58,41; 500 kişilik veri seti için 56,17; 1000 kişilik veri seti için de 58,20 değeri elde edilmiştir. Açıklanan varyans değeri için tek faktörlü ölçeklerde bu değer %30 ve üzerinde olması yeterli görülmektedir. Açıklanan varyansın yüksek olması ilgili yapı veya kavramın iyi ölçüldüğünün bir göstergesidir (Büyüköztürk, 2012). 200, 500 ve 1000 kişilik veri seti için açıklanan varyans değerleri de analizler için yeterli düzeydedir. Özdeğerler incelendiğinde ise 200 kişilik veri seti için 5,26; 500 kişilik veri seti için 5,06; 1000 kişilik veri seti için 5,24 değeri elde edilmiştir. Faktör analizinde özdeğeri 1 ya da 1'in üzerinde olan faktörler önemli kabul edilmektedir (Büyüköztürk, 2012). Bundan dolayı tüm tam veri setleri için özdeğerler de analizler için uygun düzeydedir.

Araştırmanın üçüncü alt problemünde, farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçlarının referans gruptan elde edilen sonuçlardan farklılık gösterip göstermediği betimsel düzeyde ele alınmıştır. Açımlayıcı faktör

analizinde karşılaştırma amacıyla madde faktör yük değerleri, açıklanan varyans ve özedeğerler dikkate alınmıştır.

### 3.5. Çalışma Kapsamında Ele Alınan Kayıp Veri Yöntemleri

Betimsel istatistik ve Litte'in TROK testi sonuçları incelendikten sonra veri gruplarına kayıp veri yöntemleri uygulanmıştır. Yöntem karşılaştırması için silme yöntemlerinden liste bazında silme yöntemi ve atama yöntemlerinden yaklaşık değer atama yöntemleri (seri ortalaması, yakın noktaların ortalaması, yakın noktaların medyanı, doğrusal değer kestirimi, noktanın doğrusal eğimi), beklenti maksimizasyonu, regresyon ataması ve çoklu atama yöntemleri kullanılmıştır.

#### 1. Silme Yöntemleri

- a. Liste bazında silme (LBS): Yöntemde bir ya da daha fazla kayıp veri içeren bireyler veya durumlar listeden çıkartılarak sadece tam veri içeren durumlar kullanılır (Cheema, 2012; Yılmaz, 2014)

#### 2. Atama yöntemleri

##### a. Yaklaşık değer atama yöntemleri

1. Seri ortalaması (SO): Tüm deneklerin belirli bir değişkene ilişkin ortalamasını atar.
2. Yakın noktaların ortalaması (YNO): Kayıp verinin yakınındaki değerlerin ortalamasını alarak kayıp veri yerine atama yapar.
3. Yakın noktaların medyanı (YNM): Kayıp verinin yakınındaki tam verilerin medyanını alarak kayıp veri yerine atama yapar.
4. Doğrusal değer kestirimi (DDK): Eksik veriden önceki son tam gözlem ile eksik veriden sonraki ilk tam gözlem değerini eksik veri yerine atar.
5. Noktanın doğrusal eğimi (NDE): Kayıp veri yerine mevcut yapının sahip olduğu eğilim ile uyumlu olarak bir değer atanır (Çokluk ve Kayrı, 2011).

- b. Beklenti maksimizasyonu (BM): Kayıp verileri, en çok olabilirlik kestirimleri ile dolduran iki aşamalı yöntemdir. İlk aşama olan beklenti aşamasında, kayıp veriler beklenen değerlerle tamamlanır. İkinci aşama olan maksimizasyon aşamasında ise, ilk aşamada tahmin edilen değerler kullanılarak parametre tahmini yapılır (Enders, 2001(b))
- c. Regresyon ataması (RA): Tam veriler kullanılarak, regresyon modeli elde edilir ve kayıp veriler yerine atama yapılır (Yılmaz, 2014).
- d. Çoklu atama (ÇA): Kayıp veri yerine m tane atamanın yapıldığı tekniktir. Atama sayısı genelde 3-10 arasında değişmektedir ve yöntem atfetme, analiz etme ve biraraya getirme adımlarından oluşmaktadır (Oğuzlar, 2001).

## BÖLÜM IV

### 4. Bulgular

#### 4.1. Birinci Alt Probleme İlişkin Bulgular

Normal dağılıma ve tek faktörlü yapıya sahip; 200, 500, 1000 kişilik; %5, %10 ve %20 oranlarında kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleri ile elde edilen betimsel istatistik değerleri ile tam veri setinden elde edilen betimsel istatistik değerleri arasında farklılık bulunup bulunmadığına ilişkin sonuçlar ile ortalamalar arasında manidar farklılık bulunup bulunmadığına ilişkin t-testi sonuçları Tablo 4.1., Tablo 4.2., Tablo 4.3., Tablo 4.4., Tablo 4.5 ve Tablo 4.6'da verilmiştir.

200 kişilik grup için tam veri seti ve farklı oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleriyle elde edilen betimsel istatistik değerleri Tablo 4.1'de görüldüğü gibidir.

**Tablo 4.1.** 200 kişilik veri seti için betimsel istatistik sonuçları

Kayıp veri oranı	Kayıp veri yöntemi	Ortalama ( $\bar{X}$ )	Standart sapma (S)	Çarpıklık	Basıklık
%0		18,660	5,719	,313	,214
	Seri ortalaması	18,672	5,698	,324	,225
	Yakın noktaların ortalaması	18,677	5,700	,323	,226
	Yakın noktaların medyanı	18,675	5,697	,325	,231
%5	Doğrusal değer kestirimi	18,685	5,693	,324	,231
	Noktanın doğrusal eğimi	18,672	5,700	,324	,225
	Liste bazında silme	18,721	5,732	,339	,241
	Beklenti maksimizasyonu	18,668	5,722	,313	,208
%10	Regresyon ataması	18,673	5,731	,311	,198
	Çoklu atama	18,668	5,723	,312	,209
	Seri ortalaması	18,640	5,692	,323	,271
	Yakın noktaların ortalaması	18,646	5,693	,326	,272
%10	Yakın noktaların medyanı	18,642	5,693	,326	,270
	Doğrusal değer kestirimi	18,637	5,686	,329	,283
	Noktanın doğrusal eğimi	18,640	5,691	,324	,272
	Liste bazında silme	18,666	5,833	,324	,232
%20	Beklenti maksimizasyonu	18,639	5,722	,314	,228
	Regresyon ataması	18,647	5,715	,315	,236
	Çoklu atama	18,637	5,721	,318	,231
	Seri ortalaması	18,688	5,618	,361	,282
%20	Yakın noktaların ortalaması	18,697	5,611	,356	,288
	Yakın noktaların medyanı	18,685	5,613	,357	,291
	Doğrusal değer kestirimi	18,695	5,618	,355	,289
	Noktanın doğrusal eğimi	18,686	5,618	,361	,283
%20	Liste bazında silme	18,831	5,774	,442	,313
	Beklenti maksimizasyonu	18,675	5,718	,312	,202
	Regresyon ataması	18,674	5,719	,314	,201
	Çoklu atama	18,677	5,720	,309	,203

Tablo 4.1’de görüldüğü üzere, 200 kişilik tam veri seti için hesaplanan ortalama 18,660, standart sapma 5,719, çarpıklık ,313 ve basıklık değeri ise ,214’tür. Farklı kayıp veri yöntemlerinden elde edilen değerler tam veri setinden elde edilen bu değerler dikkate alınarak karşılaştırılmıştır.

Tam veri seti için hesaplanan ortalama değeri 18,660’tır. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen ortalama değerleri incelendiğinde bu değerlerin 18,668 ile 18,721 arasında değiştiği görülmektedir. Farklı kayıp veri yöntemlerinin her biri için elde edilen ortalama değerleri tam veri setinden elde edilenden daha yüksektir. Tam veri setine göre en farklı ortalama değeri LBS yönteminde ( $\bar{X}$ =18,721) elde edilmiştir. Tam veri setine en yakın sonucu ise BM ve ÇA ( $\bar{X}$ =18,668) yöntemleri vermiştir. Farklı kayıp veri yöntemlerinden %10 kayıp veri seti için hesaplanan ortalama değerlerinin 18,637 ile 18,666 arasında değiştiği

görülmektedir. LBS yöntemi ile elde edilen ortalama ( $\bar{x}=18,666$ ) tam veri seti için hesaplanan ortalamaya en yakın sonucu verirken en farklı sonuçları DDK ve ÇA ( $\bar{x}=18,637$ ) yöntemleri vermiştir. Farklı kayıp veri yöntemlerinden %20 kayıp veri seti için hesaplanan ortalama değerleri ise 18,674 ile 18,831 arasında değişmektedir. Farklı kayıp veri yöntemlerinin her biri için elde edilen ortalama değerleri tam veri setinden elde edilenden daha yüksektir. Tam veri setine göre en farklı ortalama değeri LBS yönteminden ( $\bar{x}=18,831$ ) elde edilirken, tam veri setine en yakın değeri RA ( $\bar{x}=18,674$ ) yöntemi vermiştir.

Tam veri seti için hesaplanan standart sapma değeri 5,719'dur. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen standart sapma değerleri 5,693 ile 5,732 arasında değişen değerler almıştır. Tam veri setinden elde edilen değere en yakın değeri BM ( $S= 5,722$ ) yöntemi verirken, en farklı değeri DDK ( $S=5,693$ ) yöntemi vermiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen standart sapma değerleri 5,686 ile 5,833 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri ÇA ( $S=5,721$ ) yöntemi verirken, en farklı değeri LBS ( $S=5,833$ ) yöntemi vermiştir. Farklı kayıp veri yöntemlerinden %20 kayıp veri seti için hesaplanan standart sapma değerleri ise 5,611 ile 5,774 arasında değişen değerler almıştır. RA ( $S=5,719$ ) yöntemi tam veri setinden elde edilen değerle aynı sonucu verirken, en farklı değeri YNO ( $S=5,611$ ) yöntemi vermiştir.

Tam veri seti için hesaplanan çarpıklık değeri ,313'tür. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ,311 ile ,339 arasında değişmektedir. BM yöntemi tam veri seti ile elde edilen değer ile aynı sonucu verirken tam veri setine göre en farklı değeri LBS (Çarpıklık= ,339) yöntemi vermiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ,314 ile ,329 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri BM (Çarpıklık= ,314) yöntemi verirken, en farklı değer DDK (Çarpıklık=,329) yönteminden elde edilmiştir. %20 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ise ,309 ile ,442 arasında değişmektedir. RA yöntemi (Çarpıklık= ,314) tam veri seti ile elde edilen değere en yakın sonucu verirken tam veri setine göre en farklı değeri LBS (Çarpıklık= ,442) yöntemi vermiştir.

Tam veri seti için hesaplanan basıklık değeri ,214'tür. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ,198 ile ,241 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri ÇA (Basıklık= ,209) yöntemi verirken, en farklı değer LBS (Basıklık= ,241) yönteminden elde edilmiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ,228 ile ,283 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri BM (Basıklık= ,228) yöntemi verirken, en farklı değer DDK (Basıklık=,283) yönteminden elde edilmiştir. %20 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ise ,201 ile ,313 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri ÇA (Basıklık= ,203) yöntemi verirken, en farklı değer LBS (Basıklık= ,313) yönteminden elde edilmiştir.

200 kişilik tam veri seti için hesaplanan ortalama değeri ile %5, %10 ve %20 kayıp veri içeren veri setlerine uygulanan farklı kayıp veri yöntemlerine göre hesaplanan ortalama değerleri arasında manidar bir farklılığın bulunup bulunmadığı ilişkili örneklem t-testi ile test edilmiş ve sonuçlar Tablo 4.2'de verilmiştir.

**Tablo 4.2.** 200 kişilik veri seti için t-testi sonuçları

Kayıp veri oranı	Kayıp veri yöntemi	N	$\bar{X}$	S	sd	t	p
%5	Seri ortalaması	200	18,672	5,698	199	1,194	,234
	Tam veri		18,660	5,719			
	Yakın noktaların ortalaması		18,677	5,700		1,376	,170
	Tam veri		18,660	5,719			
	Yakın noktaların medyanı		18,675	5,697		1,096	,274
	Tam veri		18,660	5,719			
	Doğrusal değer kestirimi		18,685	5,693		1,587	,114
	Tam veri		18,660	5,719			
	Noktanın doğrusal eğimi		18,672	5,700		1,205	,229
	Tam veri		18,660	5,719			
	Liste bazında silme		18,721	5,732		-	-
	Tam veri		18,660	5,719		1,065	,288
	Beklenti maksimizasyonu		18,668	5,722			
	Tam veri		18,660	5,719			
Regresyon ataması	18,673	5,731	1,303	,194			
Tam veri	18,660	5,719					
Çoklu atama	18,668	5,723	,713	,516			
Tam veri	18,660	5,719					
%10	Seri ortalaması	200	18,640	5,692	199	-,981	,328
	Tam veri		18,660	5,719			
	Yakın noktaların ortalaması		18,646	5,693		-,662	,508
	Tam veri		18,660	5,719			
	Yakın noktaların medyanı		18,642	5,693		-,868	,387
	Tam veri		18,660	5,719			
	Doğrusal değer kestirimi		18,637	5,686		-1,000	,319
	Tam veri		18,660	5,719			
	Noktanın doğrusal eğimi		18,640	5,691		-,990	,324
	Tam veri		18,660	5,719			
	Liste bazında silme		18,666	5,833		-	-
	Tam veri		18,660	5,719		-1,262	,208
	Beklenti maksimizasyonu		18,639	5,722			
	Tam veri		18,660	5,719			
Regresyon ataması	18,647	5,715	-,706	,481			
Tam veri	18,660	5,719					
Çoklu atama	18,637	5,721	-1,107	,276			
Tam veri	18,660	5,719					
%20	Seri ortalaması	200	18,688	5,618	199	1,094	,275
	Tam veri		18,660	5,719			
	Yakın noktaların ortalaması		18,697	5,611		1,334	,184
	Tam veri		18,660	5,719			
	Yakın noktaların medyanı		18,685	5,613		,845	,399
	Tam veri		18,660	5,719			
	Doğrusal değer kestirimi		18,695	5,618		1,211	,227
	Tam veri		18,660	5,719			
	Noktanın doğrusal eğimi		18,686	5,618		1,062	,289
	Tam veri		18,660	5,719			
	Liste bazında silme		18,831	5,774		-	-
	Tam veri		18,660	5,719		,910	,364
	Beklenti maksimizasyonu		18,675	5,718			
	Tam veri		18,660	5,719			
Regresyon ataması	18,674	5,719	,709	,479			
Tam veri	18,660	5,719					
Çoklu atama	18,677	5,720	,727	,493			
Tam veri	18,660	5,719					

Tablo 4.2'deki değerler incelendiğinde, 200 kişilik tam veri setinden elde edilen ortalama değeri ile farklı kayıp oranları için farklı kayıp veri yöntemleri ile elde edilen ortalama değerleri arasında 0,05 manidarlık düzeyinde manidar bir farklılığın bulunmadığı görülmektedir. LBS yönteminde t-testine dair sonuçların raporlaştırılmamasının nedeni ise tam veri seti ile LBS yönteminin uygulandığı veri seti arasındaki standart hata farkının 0 olmasıdır.

500 kişilik grup için tam veri seti ve değişik oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleriyle elde edilen betimsel istatistik değerleri Tablo 4.3'te görüldüğü gibidir.

**Tablo 4.3.** 500 kişilik veri seti için betimsel istatistik sonuçları

Kayıp veri oranı	Kayıp veri yöntemi	Ortalama ( $\bar{X}$ )	Standart sapma (S)	Çarpıklık	Basıklık
%0	Seri ortalaması	18,724	5,600	,571	,590
	Yakın noktaların ortalaması	18,729	5,585	,579	,606
	Yakın noktaların medyanı	18,726	5,587	,577	,605
	Doğrusal değer kestirimi	18,720	5,587	,578	,608
	Doğrusal değer kestirimi	18,723	5,588	,578	,602
%5	Noktanın doğrusal eğimi	18,729	5,585	,579	,607
	Liste bazında silme	18,768	5,643	,586	,587
	Beklenti maksimizasyonu	18,725	5,602	,572	,583
	Regresyon ataması	18,712	5,610	,566	,585
	Çoklu atama	18,725	5,602	,571	,584
	Seri ortalaması	18,714	5,532	,563	,572
	Yakın noktaların ortalaması	18,719	5,520	,559	,574
	Yakın noktaların medyanı	18,716	5,523	,561	,580
%10	Doğrusal değer kestirimi	18,714	5,516	,560	,567
	Noktanın doğrusal eğimi	18,714	5,533	,563	,572
	Liste bazında silme	18,646	5,573	,572	,556
	Beklenti maksimizasyonu	18,719	5,589	,570	,567
	Regresyon ataması	18,725	5,591	,577	,585
	Çoklu atama	18,723	5,596	,565	,567
	Seri ortalaması	18,719	5,485	,589	,672
	Yakın noktaların ortalaması	18,725	5,485	,590	,676
%20	Yakın noktaların medyanı	18,720	5,485	,591	,682
	Doğrusal değer kestirimi	18,747	5,492	,590	,658
	Noktanın doğrusal eğimi	18,721	5,486	,588	,671
	Liste bazında silme	18,735	5,643	,634	,728
	Beklenti maksimizasyonu	18,731	5,590	,568	,579
	Regresyon ataması	18,741	5,584	,557	,570
	Çoklu atama	18,734	5,600	,568	,587

Tablo 4.3'te görüldüğü üzere, 500 kişilik tam veri seti için hesaplanan ortalama 18,724, standart sapma 5,600, çarpıklık ,571 ve basıklık değeri ise ,590'dır. Farklı kayıp veri yöntemlerinden elde edilen değerler tam veri setinden elde edilen bu değerler dikkate alınarak karşılaştırılmıştır.

Tam veri seti için hesaplanan ortalama değeri 18,724'tür. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen ortalama değerleri incelendiğinde bu değerlerin 18,712 ile 18,768 arasında değiştiği görülmektedir. Tam veri setinden elde edilen değere en yakın değeri BM ve ÇA ( $\bar{x}$ =18,725) yöntemleri verirken en farklı değeri LBS ( $\bar{x}$ =18,768) yöntemi vermiştir. Farklı kayıp veri yöntemlerinden %10 kayıp veri seti için hesaplanan ortalama değerlerinin 18,646 ile 18,725 arasında değiştiği görülmektedir. ÇA ( $\bar{x}$ =18,723) ve RA ( $\bar{x}$ =18,725) yöntemleri ile elde edilen ortalama tam veri seti için hesaplanan ortalamaya en yakın sonucu verirken, en farklı sonucu LBS ( $\bar{x}$ =18,646) yöntemi vermiştir. Farklı kayıp veri yöntemlerinden %20 kayıp veri seti için hesaplanan ortalama değerleri ise 18,717 ile 18,747 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri YNO ( $\bar{x}$ =18,725) yöntemi verirken en farklı değeri DDK ( $\bar{x}$ =18,747) yöntemi vermiştir.

Tam veri seti için hesaplanan standart sapma değeri 5,600'dür. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen standart sapma değerleri 5,585 ile 5,643 arasında değişen değerler almıştır. Tam veri setinden elde edilen değere en yakın değeri BM ve ÇA (S= 5,602) yöntemleri verirken, en farklı değeri LBS (S=5,643) yöntemi vermiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen standart sapma değerleri 5,516 ile 5,596 arasında değişmektedir. Farklı kayıp veri yöntemlerinin her biri için elde edilen standart sapma değerleri tam veri setinden elde edilenden daha düşüktür. Tam veri setine göre en farklı değer DDK (S=5,516) yönteminden elde edilmiştir. Tam veri setine en yakın sonucu ise ÇA (S=5,596) yöntemi vermiştir. %20 kayıp veri seti için hesaplanan standart sapma değerleri ise 5,485 ile 5,643 arasında değişen değerler almıştır. ÇA (S=5,600) yöntemi tam veri setinden elde edilen değer ile aynı sonucu verirken, tam veri setine göre en farklı değeri SO, YNO ve YNM (S=5,485) yöntemleri vermiştir.

Tam veri seti için hesaplanan çarpıklık değeri ,571'dir. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ,566 ile ,586 arasında değişmektedir. ÇA yöntemi tam veri seti ile elde edilen değer ile aynı sonucu verirken tam veri setine göre en farklı değeri LBS (Çarpıklık= ,586) yöntemi vermiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ,559 ile ,577 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri LBS (Çarpıklık= ,572) ve BM (Çarpıklık= ,570) yöntemleri verirken, en farklı değer YNO (Çarpıklık=,559) yönteminden elde edilmiştir. %20 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ise ,557 ile ,634 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri BM ve ÇA (Çarpıklık= ,568) yöntemleri verirken en farklı değer LBS (Çarpıklık= ,634) yönteminden elde edilmiştir.

Tam veri seti için hesaplanan basıklık değeri ,590'dır. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ,583 ile ,608 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri LBS (Basıklık= ,587) yöntemi verirken, en farklı değer YNM (Basıklık= ,608) yönteminden elde edilmiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ,556 ile ,585 arasında değişmektedir. Farklı kayıp veri yöntemlerinin her biri için elde edilen standart sapma değerleri tam veri setinden elde edilenden daha düşüktür. Tam veri setinden elde edilen değere en yakın değeri RA (Basıklık= ,585) yöntemi verirken, en farklı değer LBS (Basıklık=,556) yönteminden elde edilmiştir. %20 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ise ,570 ile ,728 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri ÇA (Basıklık= ,587) yöntemi verirken, en farklı değer LBS (Basıklık= ,728) yönteminden elde edilmiştir.

500 kişilik tam veri seti için hesaplanan ortalama değeri ile %5, %10 ve %20 oranlarında kayıp veri içeren veri setleri için farklı kayıp veri yöntemlerine göre hesaplanan ortalama değerleri arasında manidar bir farklılığın bulunup bulunmadığı ilişkili örneklem t-testi ile test edilmiş ve sonuçlar Tablo 4.4'te verilmiştir.

**Tablo 4.4.** 500 kişilik veri seti için t-testi sonuçları

Kayıp veri oranı	Kayıp veri yöntemi	N	$\bar{X}$	S	sd	t	p
%5	Seri ortalaması		18,729	5,585			
	Tam veri		18,724	5,600		,732	,465
	Yakın noktaların ortalaması		18,726	5,587			
	Tam veri		18,724	5,600		,226	,822
	Yakın noktaların medyanı		18,720	5,587			
	Tam veri		18,724	5,600		-,465	,642
	Doğrusal değer kestirimi		18,723	5,588			
	Tam veri		18,724	5,600		-,112	,911
	Noktanın doğrusal eğimi		18,729	5,585			
	Tam veri		18,724	5,600		,756	,450
	Liste bazında silme		18,768	5,643			
	Tam veri		18,724	5,600		-	-
	Beklenti maksimizasyonu		18,725	5,602			
	Tam veri		18,724	5,600		,218	,828
	Regresyon ataması		18,712	5,610			
	Tam veri		18,724	5,600		-1,664	,097
%10	Çoklu atama		18,725	5,602			
	Tam veri		18,724	5,600		,023	,692
	Seri ortalaması		18,714	5,532			
	Tam veri		18,724	5,600		-,752	,452
	Yakın noktaların ortalaması		18,719	5,520			
	Tam veri		18,724	5,600		-,316	,752
	Yakın noktaların medyanı		18,716	5,523			
	Tam veri		18,724	5,600		-,498	,619
	Doğrusal değer kestirimi		18,714	5,516			
	Tam veri		18,724	5,600		-,613	,540
	Noktanın doğrusal eğimi		18,714	5,533			
	Tam veri	500	18,724	5,600	499	-,772	,440
	Liste bazında silme		18,646	5,573			
	Tam veri		18,724	5,600		-	-
	Beklenti maksimizasyonu		18,719	5,589			
	Tam veri		18,724	5,600		-,505	,614
Regresyon ataması		18,725	5,591				
Tam veri		18,724	5,600		,139	,890	
%20	Çoklu atama		18,723	5,596			
	Tam veri		18,724	5,600		-,056	,564
	Seri ortalaması		18,719	5,485			
	Tam veri		18,724	5,600		-,252	,801
	Yakın noktaların ortalaması		18,725	5,485			
	Tam veri		18,724	5,600		,955	,340
	Yakın noktaların medyanı		18,720	5,485			
	Tam veri		18,724	5,600		-,167	,867
	Doğrusal değer kestirimi		18,747	5,492			
	Tam veri		18,724	5,600		1,083	,279
	Noktanın doğrusal eğimi		18,721	5,486			
	Tam veri		18,724	5,600		-,133	,894
	Liste bazında silme		18,735	5,643			
	Tam veri		18,724	5,600		-	-
	Beklenti maksimizasyonu		18,731	5,590			
	Tam veri		18,724	5,600		,685	,493
Regresyon ataması		18,741	5,584				
Tam veri		18,724	5,600		1,124	,261	
Çoklu atama		18,734	5,600				
Tam veri		18,724	5,600		,644	,383	

Tablo 4.4'teki deęerler incelendięinde, 500 kiřilik tam veri setinden elde edilen ortalama deęeri ile farklı kayıp oranları için farklı kayıp veri yöntemleri ile elde edilen ortalama deęerleri arasında 0,05 manidarlık düzeyinde manidar bir farklılığın bulunmadığı görülmektedir. LBS yönteminde t-testine dair sonuçların raporlaştırılmamasının nedeni ise tam veri seti ile LBS yönteminin uygulandığı veri seti arasındaki standart hata farkının 0 olmasıdır.

1000 kiřilik grup için tam veri seti ve farklı oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleriyle elde edilen betimsel istatistik deęerleri Tablo 4.5'de görüldüğü gibidir.

**Tablo 4.5.** 1000 kiřilik veri seti için betimsel istatistik sonuçları

Kayıp veri oranı	Kayıp veri yöntemi	Ortalama ( $\bar{X}$ )	Standart sapma (S)	Çarpıklık	Basıklık
%0		18,579	5,732	,458	,321
	Seri ortalaması	18,580	5,705	,460	,349
	Yakın noktaların ortalaması	18,580	5,711	,460	,345
	Yakın noktaların medyanı	18,576	5,713	,460	,343
%5	Doęrusal deęer kestirimi	18,580	5,715	,461	,343
	Noktanın doęrusal eğimi	18,580	5,705	,460	,349
	Liste bazında silme	18,608	5,756	,458	,340
	Beklenti maksimizasyonu	18,578	5,729	,456	,325
	Regresyon ataması	18,584	5,729	,452	,323
	Çoklu atama	18,577	5,734	,453	,320
	Seri ortalaması	18,572	5,684	,482	,369
	Yakın noktaların ortalaması	18,571	5,681	,482	,373
%10	Yakın noktaların medyanı	18,567	5,684	,481	,370
	Doęrusal deęer kestirimi	18,575	5,680	,481	,373
	Noktanın doęrusal eğimi	18,572	5,684	,481	,369
	Liste bazında silme	18,687	5,758	,479	,385
	Beklenti maksimizasyonu	18,563	5,735	,467	,321
	Regresyon ataması	18,560	5,746	,461	,313
	Çoklu atama	18,561	5,741	,465	,317
	Seri ortalaması	18,578	5,638	,452	,330
%20	Yakın noktaların ortalaması	18,574	5,642	,452	,337
	Yakın noktaların medyanı	18,567	5,643	,452	,332
	Doęrusal deęer kestirimi	18,582	5,649	,450	,333
	Noktanın doęrusal eğimi	18,578	5,639	,451	,330
	Liste bazında silme	18,435	5,836	,494	,253
	Beklenti maksimizasyonu	18,585	5,738	,453	,305
	Regresyon ataması	18,594	5,729	,453	,297
	Çoklu atama	18,592	5,742	,455	,300

Tablo 4.5'te görüldüğü üzere, 1000 kişilik tam veri seti için hesaplanan ortalama 18,579, standart sapma 5,732, çarpıklık ,458 ve basıklık değeri ise ,321'dir. Farklı kayıp veri yöntemlerinden elde edilen değerler tam veri setinden elde edilen bu değerler dikkate alınarak karşılaştırılmıştır.

Tam veri seti için hesaplanan ortalama değeri 18,579'dur. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen ortalama değerleri incelendiğinde bu değerlerin 18,576 ile 18,608 arasında değiştiği görülmektedir. Tam veri setine göre en farklı ortalama değeri LBS yönteminde ( $\bar{x}=18,608$ ) elde edilmiştir. Tam veri setine en yakın sonucu ise SO, YNO, DDK, NDE ( $\bar{x}=18,580$ ) ve BM ( $\bar{x}=18,578$ ) yöntemleri vermiştir. Farklı kayıp veri yöntemlerinden %10 kayıp veri seti için hesaplanan ortalama değerlerinin 18,560 ile 18,687 arasında değiştiği görülmektedir. DDK yöntemi ile elde edilen ortalama ( $\bar{x}=18,575$ ) tam veri seti için hesaplanan ortalamaya en yakın sonucu verirken, en farklı sonucu RA ( $\bar{x}=18,560$ ) yöntemi vermiştir. Farklı kayıp veri yöntemlerinden %20 kayıp veri seti için hesaplanan ortalama değerleri ise 18,435 ile 18,594 arasında değişmektedir. Tam veri setine göre en farklı ortalama değeri LBS yönteminden ( $\bar{x}=18,435$ ) elde edilirken, tam veri setine en yakın değeri SO ve NDE yöntemleri ( $\bar{x}=18,578$ ) vermiştir.

Tam veri seti için hesaplanan standart sapma değeri 5,732'dir. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen standart sapma değerleri 5,705 ile 5,756 arasında değişen değerler almıştır. Tam veri setinden elde edilen değere en yakın değeri ÇA ( $S= 5,734$ ) yöntemi verirken, en farklı değeri SO ve NDE ( $S=5,705$ ) yöntemleri vermiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden elde edilen standart sapma değerleri 5, 680 ile 5,758 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri BM yöntemi ( $S=5,735$ ) verirken, en farklı değeri DDK ( $S=5,680$ ) yöntemi vermiştir. Farklı kayıp veri yöntemlerinden %20 kayıp veri seti için hesaplanan standart sapma değerleri ise 5,638 ile 5,836 arasında değişen değerler almıştır. Tam veri setinden elde edilen değere en yakın değeri RA ( $S=5,729$ ) yöntemi verirken, en farklı değeri LBS ( $S=5,836$ ) yöntemi vermiştir.

Tam veri seti için hesaplanan çarpıklık değeri ,458'dir. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ,452 ile ,461 arasında değişmektedir. LBS yöntemi tam veri seti ile elde edilen değer ile aynı sonucu verirken tam veri setine göre en farklı değeri RA (Çarpıklık= ,452) yöntemi vermiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ,461 ile ,482 arasında değişmektedir. Farklı kayıp veri yöntemlerinin her biri için elde edilen çarpıklık değerleri tam veri setinden elde edilenden daha yüksektir Tam veri setinden elde edilen değere en yakın değeri RA (Çarpıklık= ,461) yöntemi verirken, en farklı değer SO ve YNO yöntemlerinden (Çarpıklık=,482) elde edilmiştir. %20 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan çarpıklık değerleri ise ,450 ile ,494 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri ÇA (Çarpıklık= ,455) yöntemi verirken, en farklı değer LBS (Çarpıklık= ,494) yönteminden elde edilmiştir.

Tam veri seti için hesaplanan basıklık değeri ,321'dir. %5 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ,320 ile ,349 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri ÇA (Basıklık= ,320) yöntemi verirken, en farklı değer SO ve NDE (Basıklık= ,349) yöntemlerinden elde edilmiştir. %10 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ,313 ile ,385 arasında değişmektedir. BM yöntemi (Basıklık= ,321) tam veri setinden elde edilen değerle aynı sonucu verirken, en farklı değer LBS (Basıklık=,385) yönteminden elde edilmiştir. %20 kayıp veri seti için farklı kayıp veri yöntemlerinden hesaplanan basıklık değerleri ise ,253 ile ,337 arasında değişmektedir. Tam veri setinden elde edilen değere en yakın değeri NDE (Basıklık= ,330) yöntemi verirken, en farklı değer LBS (Basıklık= ,253) yönteminden elde edilmiştir.

1000 kişilik tam veri seti için hesaplanan ortalama değeri ile %5, %10 ve %20 kayıp veri içeren veri setleri için farklı kayıp veri yöntemlerine göre hesaplanan ortalama değerleri arasında manidar bir farklılığın bulunup bulunmadığı ilişkili örneklem t-testi ile test edilmiş ve sonuçlar Tablo 4.6'da verilmiştir.

**Tablo 4.6.** 1000 kişilik veri seti için t-testi sonuçları

Kayıp veri oranı	Kayıp veri yöntemi	N	$\bar{X}$	S	sd	t	p																																																																																																																																																																																																																				
%5	Seri ortalaması	1000	18,580	5,705	999	,315	,753																																																																																																																																																																																																																				
	Tam veri		18,579	5,732				Yakın noktaların ortalaması	18,580	5,711	,039	,969	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,576	5,713	-,376	,707	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,580	5,715	,219	,826	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,580	5,705	,255	,789	Tam veri	18,579	5,732	Liste bazında silme	18,608	5,756	-	-	Tam veri	18,579	5,732	-,002	,998	Beklenti maksimizasyonu	18,578	5,729	Tam veri	18,579	5,732	Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732	Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,571	5,681	-,700	,484	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,684	-1,103	,270	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,575	5,680	-,325	,745	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717	,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,574	5,642	-,393	,694	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri
	Yakın noktaların ortalaması		18,580	5,711		,039	,969																																																																																																																																																																																																																				
	Tam veri		18,579	5,732				Yakın noktaların medyanı	18,576	5,713	-,376	,707	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,580	5,715	,219	,826	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,580	5,705	,255	,789	Tam veri	18,579	5,732	Liste bazında silme	18,608	5,756	-	-	Tam veri	18,579	5,732	-,002	,998	Beklenti maksimizasyonu	18,578	5,729	Tam veri	18,579	5,732	Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732	Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,571	5,681	-,700	,484	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,684	-1,103	,270	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,575	5,680	-,325	,745	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717	,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638		999		-,094	,926		Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,574	5,642	-,393	,694	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732
	Yakın noktaların medyanı		18,576	5,713		-,376	,707																																																																																																																																																																																																																				
	Tam veri		18,579	5,732				Doğrusal değer kestirimi	18,580	5,715	,219	,826	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,580	5,705	,255	,789	Tam veri	18,579	5,732	Liste bazında silme	18,608	5,756	-	-	Tam veri	18,579	5,732	-,002	,998	Beklenti maksimizasyonu	18,578	5,729	Tam veri	18,579	5,732	Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732	Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,571		5,681		-,700	,484		Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,684	-1,103	,270	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,575	5,680	-,325	,745	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717	,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638		999		-,094	,926				Tam veri	18,579		5,732	Yakın noktaların ortalaması	18,574	5,642	-,393	,694	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732		
	Doğrusal değer kestirimi		18,580	5,715		,219	,826																																																																																																																																																																																																																				
	Tam veri		18,579	5,732				Noktanın doğrusal eğimi	18,580	5,705	,255	,789	Tam veri	18,579	5,732	Liste bazında silme	18,608	5,756	-	-	Tam veri	18,579	5,732	-,002	,998	Beklenti maksimizasyonu	18,578	5,729	Tam veri	18,579	5,732	Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732	Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,571		5,681		-,700	,484		Tam veri	18,579		5,732		Yakın noktaların medyanı	18,567		5,684	-1,103	,270	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,575	5,680	-,325	,745	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717	,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638		999		-,094	,926				Tam veri	18,579				5,732	Yakın noktaların ortalaması		18,574	5,642	-,393	,694	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732				
	Noktanın doğrusal eğimi		18,580	5,705		,255	,789																																																																																																																																																																																																																				
	Tam veri		18,579	5,732				Liste bazında silme	18,608	5,756	-	-	Tam veri	18,579	5,732	-,002	,998	Beklenti maksimizasyonu	18,578	5,729	Tam veri	18,579	5,732	Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732	Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,571		5,681		-,700	,484		Tam veri	18,579		5,732		Yakın noktaların medyanı	18,567		5,684	-1,103		,270		Tam veri	18,579		5,732	Doğrusal değer kestirimi	18,575	5,680	-,325	,745	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717	,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638		999		-,094	,926				Tam veri	18,579				5,732	Yakın noktaların ortalaması				18,574	5,642		-,393	,694	Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732						
	Liste bazında silme		18,608	5,756		-	-																																																																																																																																																																																																																				
	Tam veri		18,579	5,732		-,002	,998																																																																																																																																																																																																																				
	Beklenti maksimizasyonu		18,578	5,729				Tam veri	18,579	5,732	Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732	Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,571		5,681		-,700	,484		Tam veri	18,579		5,732		Yakın noktaların medyanı	18,567		5,684	-1,103		,270		Tam veri	18,579		5,732	Doğrusal değer kestirimi		18,575		5,680	-,325		,745	Tam veri		18,579		5,732	Noktanın doğrusal eğimi		18,572	5,684	-,717	,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579		5,732		Yakın noktaların ortalaması	18,574				5,642	-,393				,694	Tam veri				18,579	5,732				Yakın noktaların medyanı	18,567		5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732													
	Tam veri		18,579	5,732		Regresyon ataması	18,584	5,729	,899	,369	Tam veri	18,579	5,732			Çoklu atama	18,577	5,734	-,347	,704	Tam veri			18,579	5,732	%10		Seri ortalaması		1000	18,572				5,684	999	-,744	,457	Tam veri		18,579					5,732	Yakın noktaların ortalaması		18,571		5,681	-,700		,484					Tam veri	18,579		5,732	Yakın noktaların medyanı		18,567		5,684				-1,103		,270		Tam veri	18,579		5,732	Doğrusal değer kestirimi			18,575	5,680	-,325	,745	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717			,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-			Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu			18,563	5,735	Tam veri		18,579		5,732	Regresyon ataması				18,560	5,746		-2,496		,013*	Tam veri				18,579						5,732				Çoklu atama	18,561				5,741	-1,980		,130			Tam veri	18,579	5,732	%20	Seri ortalaması	1000			18,578	5,638	999	-,094	,926	Tam veri			18,579	5,732	Yakın noktaların ortalaması	18,574	5,642	-,393	,694	Tam veri	18,579	5,732	Yakın noktaların medyanı			18,567	5,643	-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582			5,649	,228	,820	Tam veri	18,579	5,732			Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732
Regresyon ataması	18,584	5,729	,899	,369																																																																																																																																																																																																																							
Tam veri	18,579	5,732			Çoklu atama	18,577	5,734	-,347	,704	Tam veri	18,579	5,732	%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,571			5,681			-,700		,484	Tam veri	18,579		5,732	Yakın noktaların medyanı	18,567		5,684		-1,103	,270		Tam veri	18,579		5,732		Doğrusal değer kestirimi	18,575		5,680	-,325		,745		Tam veri	18,579		5,732	Noktanın doğrusal eğimi		18,572		5,684	-,717		,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732		Yakın noktaların ortalaması		18,574	5,642		-,393	,694	Tam veri	18,579		5,732		Yakın noktaların medyanı	18,567				5,643	-,886				,376	Tam veri				18,579	5,732	Doğrusal değer kestirimi		18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639		-,091		,927	Tam veri	18,579	5,732		Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																					
Çoklu atama	18,577	5,734	-,347	,704																																																																																																																																																																																																																							
Tam veri	18,579	5,732			%10	Seri ortalaması	1000	18,572	5,684	999	-,744	,457		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,571	5,681	-,700	,484	Tam veri	18,579			5,732			Yakın noktaların medyanı		18,567	5,684	-1,103		,270	Tam veri	18,579		5,732		Doğrusal değer kestirimi	18,575		5,680	-,325		,745		Tam veri	18,579		5,732	Noktanın doğrusal eğimi		18,572		5,684	-,717		,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,574	5,642	-,393	,694		Tam veri		18,579	5,732		Yakın noktaların medyanı	18,567	5,643	-,886		,376		Tam veri	18,579				5,732	Doğrusal değer kestirimi				18,582	5,649	,228		,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836		-		-	Tam veri	18,579	5,732		,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																													
%10	Seri ortalaması	1000	18,572	5,684		999		-,744	,457																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Yakın noktaların ortalaması	18,571		5,681		-,700	,484		Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567	5,684	-1,103			,270			Tam veri		18,579	5,732	Doğrusal değer kestirimi		18,575	5,680	-,325		,745		Tam veri	18,579		5,732	Noktanın doğrusal eğimi		18,572		5,684	-,717		,474	Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,574		5,642		-,393	,694		Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567		5,643		-,886	,376		Tam veri	18,579	5,732	Doğrusal değer kestirimi		18,582		5,649	,228				,820	Tam veri	18,579		5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu		18,585		5,738	Tam veri	18,579	5,732		Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																					
	Yakın noktaların ortalaması		18,571	5,681				-,700	,484																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Yakın noktaların medyanı	18,567		5,684		-1,103	,270		Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,575	5,680	-,325			,745			Tam veri		18,579	5,732	Noktanın doğrusal eğimi		18,572	5,684	-,717		,474		Tam veri	18,579		5,732	Liste bazında silme	18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732		Yakın noktaların ortalaması		18,574	5,642		-,393	,694		Tam veri		18,579	5,732		Yakın noktaların medyanı	18,567		5,643		-,886	,376		Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582		5,649		,228	,820		Tam veri	18,579	5,732	Noktanın doğrusal eğimi		18,578		5,639	-,091	,927		Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri		18,579		5,732	Çoklu atama	18,592	5,742		1,202	,252	Tam veri	18,579	5,732																																																
	Yakın noktaların medyanı		18,567	5,684				-1,103	,270																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Doğrusal değer kestirimi	18,575		5,680		-,325	,745		Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,572	5,684	-,717			,474			Tam veri		18,579	5,732	Liste bazında silme		18,687	5,758	-	-	Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732		Yakın noktaların ortalaması		18,574	5,642		-,393	,694	Tam veri	18,579	5,732		Yakın noktaların medyanı		18,567	5,643		-,886	,376		Tam veri		18,579	5,732		Doğrusal değer kestirimi	18,582		5,649		,228	,820		Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578		5,639		-,091	,927		Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																														
	Doğrusal değer kestirimi		18,575	5,680				-,325	,745																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Noktanın doğrusal eğimi	18,572		5,684		-,717	,474		Tam veri	18,579	5,732	Liste bazında silme	18,687	5,758	-			-			Tam veri		18,579	5,732	-2,360		,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,574	5,642	-,393	,694		Tam veri		18,579	5,732		Yakın noktaların medyanı	18,567	5,643	-,886	,376		Tam veri		18,579	5,732		Doğrusal değer kestirimi	18,582		5,649		,228	,820		Tam veri	18,579		5,732		Noktanın doğrusal eğimi	18,578		5,639	-,091	,927	Tam veri	18,579		5,732		Liste bazında silme	18,435		5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																			
	Noktanın doğrusal eğimi		18,572	5,684				-,717	,474																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Liste bazında silme	18,687		5,758		-	-		Tam veri	18,579	5,732	-2,360	,018*	Beklenti maksimizasyonu	18,563	5,735	Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,574	5,642		-,393		,694	Tam veri		18,579	5,732		Yakın noktaların medyanı		18,567	5,643		-,886	,376	Tam veri	18,579	5,732		Doğrusal değer kestirimi		18,582	5,649		,228	,820	Tam veri	18,579	5,732		Noktanın doğrusal eğimi		18,578	5,639		-,091	,927		Tam veri		18,579	5,732		Liste bazında silme	18,435		5,836		-	-		Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																	
	Liste bazında silme		18,687	5,758				-	-																																																																																																																																																																																																																		
	Tam veri		18,579	5,732				-2,360	,018*																																																																																																																																																																																																																		
	Beklenti maksimizasyonu		18,563	5,735							Tam veri	18,579	5,732	Regresyon ataması	18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,574	5,642	-,393	,694		Tam veri		18,579	5,732		Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732		Doğrusal değer kestirimi		18,582	5,649		,228	,820		Tam veri		18,579	5,732		Noktanın doğrusal eğimi	18,578	5,639	-,091	,927		Tam veri		18,579	5,732		Liste bazında silme	18,435	5,836	-	-		Tam veri		18,579	5,732		,835	,404		Beklenti maksimizasyonu		18,585	5,738		Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																	
	Tam veri		18,579	5,732	Regresyon ataması		18,560	5,746	-2,496	,013*	Tam veri	18,579	5,732	Çoklu atama	18,561	5,741			-1,980	,130	Tam veri	18,579	5,732	%20			Seri ortalaması	1000	18,578		5,638		999	-,094				,926	Tam veri	18,579	5,732	Yakın noktaların ortalaması	18,574				5,642		-,393	,694		Tam veri	18,579	5,732			Yakın noktaların medyanı	18,567	5,643		-,886		,376	Tam veri					18,579		5,732	Doğrusal değer kestirimi		18,582	5,649	,228				,820		Tam veri	18,579		5,732	Noktanın doğrusal eğimi	18,578	5,639	-,091		,927		Tam veri	18,579				5,732	Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu			18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması			18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																					
Regresyon ataması	18,560	5,746	-2,496	,013*																																																																																																																																																																																																																							
Tam veri	18,579	5,732			Çoklu atama	18,561	5,741	-1,980	,130	Tam veri	18,579	5,732	%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926	Tam veri	18,579	5,732		Yakın noktaların ortalaması	18,574	5,642		-,393		,694			Tam veri		18,579	5,732	Yakın noktaların medyanı	18,567	5,643	-,886	,376	Tam veri	18,579	5,732		Doğrusal değer kestirimi		18,582	5,649		,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578	5,639		-,091		,927	Tam veri		18,579	5,732		Liste bazında silme		18,435	5,836		-	-	Tam veri	18,579	5,732		,835		,404	Beklenti maksimizasyonu		18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																		
Çoklu atama	18,561	5,741	-1,980	,130																																																																																																																																																																																																																							
Tam veri	18,579	5,732			%20	Seri ortalaması	1000	18,578	5,638	999	-,094	,926		Tam veri		18,579	5,732		Yakın noktaların ortalaması	18,574	5,642	-,393	,694		Tam veri	18,579	5,732		Yakın noktaların medyanı		18,567			5,643		-,886	,376	Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582	5,649	,228	,820		Tam veri		18,579	5,732		Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732		Liste bazında silme		18,435	5,836		-	-		Tam veri		18,579	5,732		,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																										
%20	Seri ortalaması	1000	18,578	5,638		999		-,094	,926																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Yakın noktaların ortalaması	18,574		5,642		-,393	,694		Tam veri	18,579	5,732	Yakın noktaların medyanı	18,567		5,643	-,886	,376		Tam veri		18,579			5,732		Doğrusal değer kestirimi	18,582	5,649	,228	,820	Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578		5,639		-,091	,927		Tam veri	18,579	5,732	Liste bazında silme	18,435	5,836	-	-		Tam veri		18,579	5,732		,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																		
	Yakın noktaların ortalaması		18,574	5,642				-,393	,694																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Yakın noktaların medyanı	18,567		5,643		-,886	,376		Tam veri	18,579	5,732	Doğrusal değer kestirimi	18,582		5,649	,228	,820		Tam veri		18,579			5,732		Noktanın doğrusal eğimi	18,578	5,639	-,091	,927	Tam veri	18,579	5,732	Liste bazında silme	18,435		5,836		-	-		Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																													
	Yakın noktaların medyanı		18,567	5,643				-,886	,376																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Doğrusal değer kestirimi	18,582		5,649		,228	,820		Tam veri	18,579	5,732	Noktanın doğrusal eğimi	18,578		5,639	-,091	,927		Tam veri		18,579			5,732		Liste bazında silme	18,435	5,836	-	-	Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																																								
	Doğrusal değer kestirimi		18,582	5,649				,228	,820																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Noktanın doğrusal eğimi	18,578		5,639		-,091	,927		Tam veri	18,579	5,732	Liste bazında silme	18,435		5,836	-	-		Tam veri		18,579			5,732		,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																																																
	Noktanın doğrusal eğimi		18,578	5,639				-,091	,927																																																																																																																																																																																																																		
	Tam veri		18,579	5,732							Liste bazında silme	18,435		5,836		-	-		Tam veri	18,579	5,732	,835	,404	Beklenti maksimizasyonu	18,585	5,738	Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																																																														
	Liste bazında silme		18,435	5,836				-	-																																																																																																																																																																																																																		
	Tam veri		18,579	5,732				,835	,404																																																																																																																																																																																																																		
	Beklenti maksimizasyonu		18,585	5,738							Tam veri	18,579	5,732	Regresyon ataması	18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																																																																														
	Tam veri		18,579	5,732	Regresyon ataması		18,594	5,729	1,430	,153	Tam veri	18,579	5,732	Çoklu atama	18,592	5,742			1,202	,252	Tam veri	18,579	5,732																																																																																																																																																																																																				
Regresyon ataması	18,594	5,729	1,430	,153																																																																																																																																																																																																																							
Tam veri	18,579	5,732			Çoklu atama	18,592	5,742	1,202	,252	Tam veri	18,579	5,732																																																																																																																																																																																																															
Çoklu atama	18,592	5,742	1,202	,252																																																																																																																																																																																																																							
Tam veri	18,579	5,732																																																																																																																																																																																																																									

Tablo 4-6'deki deęerler incelendięinde, 1000 kiřilik tam veri setinden elde edilen ortalama deęeri ile farklı kayıp veri oranları için farklı kayıp veri yöntemleri ile elde edilen ortalama deęerleri arasında 0,05 manidarlık düzeyinde %10 kayıp durumunda BM ve RA yöntemleri hariç manidar bir farklılığın bulunmadığı görülmektedir. %10 kayıp veri durumunda BM ve RA yöntemlerinde 0,05 manidarlık düzeyinde manidar bir farklılığın olduğu görülmektedir. LBS yönteminde t-testine dair sonuçların raporlaştırılmamasının nedeni ise tam veri seti ile LBS yönteminin uygulandığı veri seti arasındaki standart hata farkının 0 olmasıdır.

#### 4.2. İkinci Alt Probleme İliřkin Bulgular

Normal dağılıma sahip, tek faktörlü 200, 500, 1000 kiřilik; %5, %10 ve %20 oranlarında kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleri ile elde edilen cronbach  $\alpha$  güvenilirlik katsayıları ve tam veri setinden elde edilen Cronbach  $\alpha$  güvenilirlik katsayısı ile bunlar arasında manidar farklılık bulunup bulunmadığına ilişkin olarak gerçekleştirilen Fisher'in z testi sonuçları Tablo 4.7, Tablo 4.8 ve Tablo 4.9'da verilmiştir.

200 kiřilik grup için tam veri seti ve farklı oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleriyle elde edilen cronbach  $\alpha$  güvenilirlik katsayıları ile Fisher'in z testi sonuçları Tablo 4.7'de görüldüğü gibidir.

**Tablo 4.7.** Farklı oranlarda kayıp veri içeren 200 kişilik veri seti için güvenilirlik katsayıları ve Fisher z testi sonuçları

Kayıp Veri Yöntemleri	%0 kayıp Cronbach $\alpha$ (Fisher z)	%5 kayıp Cronbach $\alpha$ (Fisher z)	%10 kayıp Cronbach $\alpha$ (Fisher z)	%20 kayıp Cronbach $\alpha$ (Fisher z)
Seri ortalaması		,909 (,000)	,911 (,000)	,906 (,290)
Yakın noktaların ortalaması		,909 (,000)	,910 (,000)	,905 (,290)
Yakın noktaların medyanı		,909 (,000)	,910 (,000)	,904 (,290)
Doğrusal değer kestirimi		,908 (,000)	,910 (,000)	,905 (,290)
Noktanın doğrusal eğimi	,910 (,000)	,909 (,000)	,910 (,000)	,906 (,290)
Liste bazında silme		,910 (,000)	,914 (-,290)	,913 (-,290)
Beklenti maksimizasyonu		,910 (,000)	,912 (,000)	,911 (,000)
Regresyon ataması		,911 (,000)	,911 (,000)	,910 (,000)
Çoklu atama		,910 (,000)	,911 (,000)	,909 (,000)

Tablo 4.7.'de görüldüğü üzere 200 kişilik tam veri seti için hesaplanan cronbach alpha güvenilirlik katsayısı ,910'dur. %5 oranında kayıp veri içeren 200 kişilik veri seti için farklı kayıp veri yöntemleriyle elde edilen cronbach alpha güvenilirlik katsayıları ,908 ile ,911 arasında değişmektedir. En düşük cronbach alpha katsayısı DDK (Cronbach  $\alpha=$  ,908) yönteminden elde edilirken, LBS, BM ve ÇA yöntemleri (Cronbach  $\alpha=$  ,910) tam veri seti ile aynı katsayıyı vermişlerdir. SO, YNO, YNM, NDE ve RA yöntemleri ise tam veri setine yakın katsayılar verirken, en yüksek katsayı RA (Cronbach  $\alpha=$  ,911) yönteminden elde edilmiştir.

%10 oranında kayıp veri içeren 200 kişilik veri seti için farklı kayıp veri yöntemleri kullanılarak elde edilen cronbach alpha güvenilirlik katsayılarının ,910 ile ,914 arasında değiştiği görülmektedir. SO (Cronbach  $\alpha=$  ,911) yöntemi hariç yaklaşık değer atama yöntemleri tam veri seti ile aynı katsayıyı verirken; en yüksek katsayıyı LBS (Cronbach  $\alpha=$ ,914) yöntemi vermiştir. SO, RA ve ÇA (Cronbach  $\alpha=$  ,911) yöntemleri ise tam veri setine göre yüksek katsayılar verse de bu katsayılar tam veri setinden elde edilen katsayıya oldukça yakındır.

%20 oranında kayıp veri içeren 200 kişilik veri seti için farklı kayıp veri yöntemleri kullanılarak elde edilen cronbach alpha güvenilirlik katsayılarının ise ,904 ile

,913 arasında deđiřtiđi grlmektedir. En dřk katsayıyı YNM (Cronbach  $\alpha = ,904$ ) yntemi verirken, en yksek katsayı LBS (Cronbach  $\alpha = ,913$ ) yntemi ile elde edilmiřtir. RA (Cronbach  $\alpha = ,910$ ) yntemi ile elde edilen katsayı tam veri setinden elde edilen katsayı ile aynı iken BM (Cronbach  $\alpha = ,911$ ) ve A (Cronbach  $\alpha = ,909$ ) yntemleri tam veri setinden elde edilen katsayıya en yakın katsayıları vermiřtir. Yaklařık deđer atama yntemlerinden elde edilen katsayıların ise tam veri setinden elde edilene gre dřk olduđu grlmektedir.

Tablo 4.7'deki 200 kiřilik tam veri setinden elde edilen cronbach alpha gvenirlik katsayısı ile farklı oranlarda kayıp veri ieren veri setlerinden farklı kayıp veri yntemleri ile elde edilen Cronbach alpha gvenirlik katsayıları arasında manidar bir farklılık bulunup bulunmadıđına iliřkin olarak gerekleřtirilen Fisher'in  $z$  testi sonuları incelendiđinde, tm kayıp veri oranları ve tm kayıp veri yntemleri iin hesaplanan  $z$  deđerlerinin -1,96 ile + 1,96 aralıđında deđerler aldıđı grlmektedir. Bu bulgu, tam veri seti ile kayıp veri yntemlerinin uygulandıđı veri setlerinden elde edilen Cronbach alpha gvenirlik katsayıları arasında manidar bir farklılıđın bulunmadıđını gstermektedir.

500 kiřilik grup iin tam veri seti ve farklı oranlarda kayıp veri ieren veri setlerinden farklı kayıp veri yntemleriyle elde edilen cronbach  $\alpha$  gvenirlik katsayıları ile Fisher'in  $z$  testi sonuları Tablo 4.8'de grldđ gibidir.

**Tablo 4.8.** Farklı oranlarda kayıp veri içeren 500 kişilik veri seti için güvenilirlik katsayıları ve Fisher z testi sonuçları

Kayıp Veri Yöntemleri	%0 kayıp Cronbach $\alpha$ (Fisher z)	%5 kayıp Cronbach $\alpha$ (Fisher z)	%10 kayıp Cronbach $\alpha$ (Fisher z)	%20 kayıp Cronbach $\alpha$ (Fisher z)
Seri ortalaması		,902 (,000)	,899 (,000)	,898 (,000)
Yakın noktaların ortalaması		,901 (,000)	,898 (,000)	,896 (,394)
Yakın noktaların medyanı		,901 (,000)	,898 (,000)	,896 (,394)
Doğrusal değer kestirimi		,901 (,000)	,897 (,394)	,895 (,394)
Noktanın doğrusal eğimi	,902 (,000)	,902 (,000)	,899 (,000)	,898 (,000)
Liste bazında silme		,905 (-,425)	,901 (,000)	,904 (-,425)
Beklenti maksimizasyonu		,903 (-,425)	,903 (-,425)	,903 (-,425)
Regresyon ataması		,903 (-,425)	,902 (,000)	,901 (,000)
Çoklu atama		,902 (,000)	,902 (,000)	,901 (,000)

Tablo 4.8.'de görüldüğü üzere 500 kişilik tam veri seti için hesaplanan cronbach alpha güvenilirlik katsayısı ,902'dir. %5 oranında kayıp veri içeren 500 kişilik veri seti için farklı kayıp veri yöntemleriyle elde edilen cronbach alpha güvenilirlik katsayıları ,901 ile ,905 arasında değişmektedir. En düşük cronbach alpha katsayısı YNO, YNM ve DDK (Cronbach  $\alpha$ = ,901) yöntemlerinden elde edilirken, SO, NDE ve ÇA yöntemleri (Cronbach  $\alpha$ = ,902) tam veri seti ile aynı katsayıyı vermişlerdir. En yüksek katsayı ise LBS (Cronbach  $\alpha$ = ,905) yönteminden elde edilmiştir.

%10 oranında kayıp veri içeren 500 kişilik veri seti için farklı kayıp veri yöntemleri kullanılarak elde edilen cronbach alpha güvenilirlik katsayılarının ,897 ile ,903 arasında değiştiği görülmektedir. En düşük cronbach alpha katsayısı DDK (Cronbach  $\alpha$ = ,897) yönteminden elde edilirken, RA ve ÇA yöntemleri (Cronbach  $\alpha$ = ,902) tam veri seti ile aynı katsayıyı vermiştir. En yüksek katsayı ise BM yönteminden (Cronbach  $\alpha$ =,903) elde edilmiştir. Yaklaşık değer atama yöntemlerinden elde edilen katsayıların ise tam veri setinden elde edilene göre düşük olduğu görülmektedir.

%20 oranında kayıp veri içeren 500 kişilik veri seti için farklı kayıp veri yöntemleri kullanılarak elde edilen cronbach alpha güvenilirlik katsayılarının ise ,895 ile ,904 arasında değiştiği görülmektedir. En düşük katsayıyı DDK (Cronbach  $\alpha$ = ,895) yöntemi verirken, en yüksek katsayı LBS (Cronbach  $\alpha$ = ,904) yöntemi ile elde

edilmiştir. RA, ÇA (Cronbach  $\alpha = ,901$ ) ve BM (Cronbach  $\alpha = ,903$ ) yöntemleri tam veri setinden elde edilen katsayıya en yakın katsayıları vermiştir. Yaklaşık değer atama yöntemlerinden elde edilen katsayıların ise tam veri setinden elde edilene göre düşük olduğu görülmektedir.

Tablo 4.8'deki 500 kişilik tam veri setinden elde edilen cronbach alpha güvenilirlik katsayısı ile farklı oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleri ile elde edilen Cronbach alpha güvenilirlik katsayıları arasında manidar bir farklılık bulunup bulunmadığına ilişkin olarak gerçekleştirilen Fisher'in  $z$  testi sonuçları incelendiğinde, tüm kayıp veri oranları ve tüm kayıp veri yöntemleri için hesaplanan  $z$  değerlerinin  $-1,96$  ile  $+1,96$  aralığında değerler aldığı görülmektedir. Bu bulgu, tam veri seti ile kayıp veri yöntemlerinin uygulandığı veri setlerinden elde edilen Cronbach alpha güvenilirlik katsayıları arasında manidar bir farklılığın bulunmadığını göstermektedir.

1000 kişilik grup için tam veri seti ve farklı oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleriyle elde edilen cronbach  $\alpha$  güvenilirlik katsayıları ile Fisher'in  $z$  testi sonuçları Tablo 4.9'da görüldüğü gibidir.

**Tablo 4.9.** Farklı oranlarda kayıp veri içeren 1000 kişilik veri seti için güvenilirlik katsayıları ve Fisher  $z$  testi sonuçları

Kayıp Veri Yöntemleri	%0 kayıp Cronbach $\alpha$ (Fisher $z$ )	%5 kayıp Cronbach $\alpha$ (Fisher $z$ )	%10 kayıp Cronbach $\alpha$ (Fisher $z$ )	%20 kayıp Cronbach $\alpha$ (Fisher $z$ )
Seri ortalaması		,909 (,000)	,908 (,000)	,906 (,650)
Yakın noktaların ortalaması		,909 (,000)	,907 (,650)	,905 (,650)
Yakın noktaların medyanı		,909 (,000)	,907 (,650)	,904 (,650)
Doğrusal değer kestirimi		,909 (,000)	,907 (,650)	,904 (,650)
Noktanın doğrusal eğimi	,910 (,000)	,909 (,000)	,908 (,000)	,906 (,650)
Liste bazında silme		,911 (,000)	,911 (,000)	,913 (-,650)
Beklenti maksimizasyonu		,910 (,000)	,911 (,000)	,911 (,000)
Regresyon ataması		,909 (,000)	,910 (,000)	,909 (,000)
Çoklu atama		,910 (,000)	,910 (,000)	,910 (,000)

Tablo 4.9.'da görüldüğü üzere 1000 kişilik tam veri seti için hesaplanan cronbach alpha güvenilirlik katsayısı ,910'dur. %5 oranında kayıp veri içeren 1000 kişilik veri seti için farklı kayıp veri yöntemleriyle elde edilen cronbach alpha güvenilirlik katsayıları ,909 ile ,911 arasında değişmektedir. En düşük cronbach alpha katsayısı yaklaşık değer atama yöntemleri ve RA (Cronbach  $\alpha=$  ,909) yöntemlerinden elde edilirken, BM ve ÇA yöntemleri (Cronbach  $\alpha=$  ,910) tam veri seti ile aynı katsayıyı vermişlerdir. Tüm yöntemler tam veri setine yakın değerler verirken, en yüksek katsayı LBS (Cronbach  $\alpha=$  ,911) yönteminden elde edilmiştir. Ancak bu katsayı da tam veri setinden elde edilen katsayıya oldukça yakındır.

%10 oranında kayıp veri içeren 1000 kişilik veri seti için farklı kayıp veri yöntemleri kullanılarak elde edilen cronbach alpha güvenilirlik katsayılarının ,907 ile ,911 arasında değiştiği görülmektedir. En düşük cronbach alpha katsayısı YNO, YNM ve DDK yöntemlerinden (Cronbach  $\alpha=$  ,907) elde edilirken, RA ve ÇA (Cronbach  $\alpha=$  ,910) yöntemleri tam veri seti ile aynı katsayıyı vermiştir. En yüksek katsayıyı ise LBS ve BM (Cronbach  $\alpha=$ ,911) yöntemleri vermiştir. SO ve NDE (Cronbach  $\alpha=$  ,908) yöntemleri tam veri setine göre düşük katsayı verse de bu katsayılar tam veri setinden elde edilen değere oldukça yakındır.

%20 oranında kayıp veri içeren 1000 kişilik veri seti için farklı kayıp veri yöntemleri kullanılarak elde edilen cronbach alpha güvenilirlik katsayılarının ise ,904 ile ,913 arasında değiştiği görülmektedir. En düşük katsayıyı YNM ve DDK (Cronbach  $\alpha=$  ,904) yöntemleri verirken, en yüksek katsayı LBS (Cronbach  $\alpha=$  ,913) yöntemi ile elde edilmiştir. ÇA (Cronbach  $\alpha=$  ,910) yöntemi ile elde edilen katsayı tam veri setinden elde edilen katsayı ile aynı iken BM (Cronbach  $\alpha=$  ,911) ve RA (Cronbach  $\alpha=$  ,909) yöntemleri tam veri setinden elde edilen katsayıya en yakın değerleri vermiştir. Yaklaşık değer atama yöntemlerinden elde edilen katsayıların ise tam veri setinden elde edilene göre düşük olduğu görülmektedir.

Tablo 4.9'daki 1000 kişilik tam veri setinden elde edilen cronbach alpha güvenilirlik katsayısı ile farklı oranlarda kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleri ile elde edilen Cronbach alpha güvenilirlik katsayıları arasında manidar

bir farklılık bulunup bulunmadığına ilişkin olarak gerçekleştirilen Fisher'in  $z$  testi sonuçları incelendiğinde, tüm kayıp veri oranları ve tüm kayıp veri yöntemleri için hesaplanan  $z$  değerlerinin  $-1,96$  ile  $+ 1,96$  aralığında değerler aldığı görülmektedir. Bu bulgu, tam veri seti ile kayıp veri yöntemlerinin uygulandığı veri setlerinden elde edilen Cronbach alpha güvenirlik katsayıları arasında manidar bir farklılığın bulunmadığını göstermektedir.

#### 4.3. Üçüncü Alt Probleme İlişkin Bulgular

Normal dağılıma sahip, tek faktörlü 200, 500, 1000 kişilik; %5, %10 ve %20 oranlarında kayıp veri içeren veri setlerinden farklı kayıp veri yöntemleri neticesinde elde edilen açımlayıcı faktör analizi sonuçları (madde faktör yükleri, açıklanan varyans ve özdeğerler) ile tam veri setinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4.10, Tablo 4.11, Tablo 4.12, Tablo 4.13, Tablo 4.14, Tablo 4.15, Tablo 4.16, Tablo 4.17 ve Tablo 4.18'de verilmiştir.

200 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4-10.'da görüldüğü gibidir.

**Tablo 4-10.** 200 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,71	,71	,71	,71	,70	,71	,71	,71	,72	,71
2	,78	,79	,79	,79	,79	,79	,79	,79	,79	,79
3	,77	,77	,77	,77	,77	,77	,77	,77	,77	,77
4	,73	,73	,77	,77	,72	,73	,72	,73	,73	,73
5	,80	,80	,80	,80	,80	,80	,80	,80	,80	,80
6	,81	,81	,81	,81	,80	,81	,80	,81	,81	,81
7	,77	,77	,77	,77	,77	,77	,77	,77	,77	,77
8	,75	,75	,76	,75	,75	,75	,76	,76	,76	,76
9	,75	,75	75	,75	,75	,75	,74	,75	,75	,75
Açıklanan varyans	58,41	58,21	58,17	58,07	57,93	58,22	58,31	58,54	58,60	58,46
Özdeğer	5,26	5,24	5,24	5,23	5,21	5,24	5,25	5,27	5,27	5,26

Tablo 4.10'da görüldüğü üzere, 200 kişilik tam veri seti için hesaplanan madde faktör yük değerleri ,71 ile ,81 arasında değişmekte olup bu veri seti için hesaplanan açıklanan varyans değeri 58,41 ve özdeğer ise 5,26'dır. Farklı kayıp veri yöntemlerinden elde edilen değerler tam veri setinden elde edilen bu değerler dikkate alınarak karşılaştırılmıştır.

Tablo 4.10'a göre, 200 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,71 ile ,81 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde SO ve NDE yöntemleri için 2.madde hariç tam veri seti ile aynı faktör yük değerlerinin elde edildiği görülmektedir. YNM (2. ve 4. madde), BM (2. ve 8. madde) ve ÇA (2. ve 8. madde) yöntemlerinde 2 madde hariç tam veri seti ile aynı faktör yük değerleri elde edilmiştir. YNO (2,4, ve 8.madde) ve RA (1, 2 ve 8. madde) yöntemlerinde ise 3 madde hariç tam veri seti ile aynı sonuçlar elde edilmiştir. DDK (1,2,4 ve 6.madde) yönteminde 4 maddede, LBS

(2,4,6,8 ve 9.madde) yönteminde ise 5 maddede, tam veri setinden elde edilen değerlere göre, farklı faktör yük değerleri elde edilmiştir.

Tablo 4.10'a göre, 200 kişilik tam veri seti için elde edilen açıklanan varyans değeri 58,41'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri RA (58,60) yönteminden elde edilirken, en düşük değer DDK (57,93) yönteminden elde edilmiştir. Tam veri setinden elde edilen değere en yakın hesaplanan değer ise ÇA (58,46) yönteminden elde edilmiştir.

Tablo 4.10'a göre, 200 kişilik tam veri seti için elde edilen özdeğer 5,26'dır. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,21 ile 5,27 arasında değişmektedir. ÇA yöntemi için hesaplan özdeğer tam veri seti için hesaplanan öz değerle aynıdır. En düşük ve tam veri setinden elde edilen öz değere en uzak değer DDK (5.21) yönteminden elde edilirken, en yüksek özdeğeri BM ve RA (5,27) yöntemleri vermiştir.

200 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4.11.'de görüldüğü gibidir.

**Tablo 4.11.** 200 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,71	,72	,72	,72	,71	,72	,73	,73	,72	,73
2	,78	,79	,79	,78	,79	,79	,80	,79	,79	,79
3	,77	,77	,77	,77	,77	,77	,77	,77	,78	,77
4	,73	,74	,74	,74	,74	,74	,75	,74	,74	,74
5	,78	,79	,79	,79	,79	,79	,80	,79	,79	,79
6	,81	,81	,80	,80	,80	,81	,82	,81	,81	,81
7	,77	,78	,78	,78	,78	,78	,78	,78	,78	,78
8	,75	,75	,75	,75	,75	,75	,75	,75	,75	,75
9	,75	,74	,74	,74	,74	,74	,75	,75	,74	,74
Açıklanan varyans	58,41	58,60	58,52	58,45	58,36	58,29	59,57	59,03	58,72	58,97
Özdeğer	5,26	5,27	5,27	5,26	5,25	5,27	5,36	5,31	5,29	5,30

Tablo 4.11'e göre, 200 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,71 ile ,81 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde BM (3,6,8 ve 9.madde) yönteminde 4 madde hariç tam veri seti ile farklı faktör yük değerleri elde edildiği görülmektedir. SO (3,6,8.madde), YNM (2,3,8.madde), DDK (1,3,8.madde), NDE (3,6,8.madde), LBS (3,8 ve 9.madde) ve ÇA (3,6,8.madde) yöntemlerinde 3'er madde tam veri seti ile aynı sonucu vermiştir. Tam veri seti ile en az benzerlik gösteren yöntemler ise YNO ve RA yöntemleridir ve bu yöntemler için 2 madde tam veri seti ile aynı sonucu vermiştir.

Tablo 4.11'e göre, 200 kişilik tam veri seti için elde edilen açıklanan varyans değeri 58,41'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri LBS (59,57) yönteminden elde edilirken, en düşük değer NDE (58,29)

yönteminden elde edilmiştir. Tam veri setinden elde edilen değere en yakın hesaplanan değer ise YNM (58,45) yönteminden elde edilmiştir.

Tablo 4.11'e göre, 200 kişilik tam veri seti için elde edilen özdeğer 5,26'dır. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,25 ile 5,36 arasında değişmektedir. YNM (5,26) yöntemi için hesaplan özdeğer tam veri seti için hesaplanan özdeğerle aynıdır. En düşük özdeğer DDK (5,25) yönteminden elde edilirken, en yüksek ve tam veri setine en uzak özdeğeri LBS (5,36) yöntemi vermiştir.

200 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4.12.'de görüldüğü gibidir.

**Tablo 4.12.** 200 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,71	,72	,72	,72	,72	,72	,73	,72	,73	,72
2	,78	,76	,77	,77	,76	,76	,77	,78	,77	,77
3	,77	,75	,74	,75	,75	,75	,78	,77	,77	,76
4	,73	,72	,72	,71	,71	,72	,71	,73	,73	,73
5	,80	,79	,78	,78	,78	,78	,80	,80	,79	,79
6	,81	,81	,80	,80	,81	,81	,82	,81	,81	,81
7	,77	,78	,77	,78	,77	,78	,79	,77	,77	,77
8	,75	,75	,75	,75	,75	,76	,77	,76	,76	,75
9	,75	,75	,75	,75	,74	,75	,77	,75	,75	,75
Açıklanan varyans	58,41	57,48	57,00	56,92	57,04	57,47	59,22	58,69	58,52	58,17
Özdeğer	5,26	5,17	5,13	5,12	5,13	5,17	5,33	5,28	5,27	5,24

Tablo 4.12'ye göre, 200 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,71 ile ,81 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde tam veri setine en yakın sonuçları BM yönteminin verildiği görülmektedir. BM yönteminde 2 madde hariç (1.ve 8.madde) tam veri seti ile aynı madde faktör yük değerleri elde edilmiştir. BM yönteminden sonra en yakın sonuçları RA ve ÇA yöntemleri vermiştir. RA (1,2,5 ve 8.madde) ve ÇA (1,2,3 ve 5.madde) yöntemlerinde 4 madde hariç tam veri seti ile aynı sonuç elde edilmiştir. SO (6,8 ve 9.madde), YNO (7,8 ve 9.madde), DDK (6,7 ve 8.madde) yöntemlerinde 3 madde tam veri seti ile benzerlik gösterirken; YNM (8 ve 9.madde) ve NDE (6 ve 9.madde) yöntemlerinde 2 madde tam veri seti ile benzerlik göstermektedir. LBS yönteminde ise sadece 5.maddenin tam veri seti ile aynı sonuçları verdiği görülmektedir.

Tablo 4.12'ye göre, 200 kişilik tam veri seti için elde edilen açıklanan varyans değeri 58,41'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri LBS (59,22) yönteminden elde edilirken, en düşük değer YNM (56,92) yönteminden elde edilmiştir. Tam veri setinden elde edilen değere en yakın hesaplanan değer ise RA (58,52) yönteminden elde edilmiştir.

Tablo 4.12'ye göre, 200 kişilik tam veri seti için elde edilen özdeğer 5,26'dır. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,12 ile 5,33 arasında değişmektedir. RA yöntemi için hesaplan özdeğer tam veri setine en yakın sonucu vermiştir. En düşük ve tam veri setinden elde edilen özdeğere en uzak değer YNM (5,12) yönteminden elde edilirken, en yüksek özdeğeri LBS (5,33) yöntemi vermiştir.

500 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açılımlayıcı faktör analizi sonuçları Tablo 4.13.'te görüldüğü gibidir.

**Tablo 4.13.** 500 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,75	,75	,75	,75	,75	,75	,75	,75	,75	,75
2	,79	,78	,78	,78	,78	,78	,79	,79	,79	,79
3	,74	,74	,74	,74	,74	,74	,74	,74	,74	,74
4	,72	,72	,72	,72	,72	,72	,73	,72	,72	,72
5	,78	,78	,78	,78	,78	,78	,78	,78	,78	,78
6	,77	,78	,78	,78	,78	,78	,79	,78	,78	,78
7	,72	,72	,72	,72	,72	,72	,73	,72	,72	,72
8	,74	,74	,74	,74	,74	,74	,75	,74	,74	,74
9	,74	,74	,74	,74	,74	,74	,74	,74	,74	,74
Açıklanan varyans	56,17	56,11	56,06	56,06	56,04	56,10	56,82	56,36	56,32	56,24
Özdeğer	5,06	5,06	5,05	5,05	5,04	5,05	5,11	5,07	5,07	5,06

Tablo 4.13'te görüldüğü üzere, 500 kişilik tam veri seti için hesaplanan madde faktör yük değerleri ,72 ile ,79 arasında değişmekte olup bu veri seti için hesaplanan açıklanan varyans değeri 56,17 ve özdeğer ise 5,06'dır. Farklı kayıp veri yöntemlerinden elde edilen değerler tam veri setinden elde edilen bu değerler dikkate alınarak karşılaştırılmıştır.

Tablo 4.13'e göre, 500 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,72 ile ,79 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde yaklaşık değer atama yöntemlerinde 2.ve 6.maddeler hariç tam veri seti ile aynı madde faktör yük değeri elde edildiği görülmektedir. BM, RA ve ÇA yöntemleri 6.madde hariç tam veri seti ile aynı madde faktör yük değerlerini verirken, LBS yönteminde 5 madde tam veri seti ile aynı sonucu vermiştir.

Tablo 4.13'e göre, 500 kişilik tam veri seti için elde edilen açıklanan varyans değeri 56,17'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri LBS (56,82) yönteminden elde edilirken, en düşük değer DDK (56,04) yönteminden elde edilmiştir. Tam veri setinden elde edilen değere en yakın hesaplanan değer ise SO (56,11) yönteminden elde edilmiştir.

Tablo 4.13'e göre, 500 kişilik tam veri seti için elde edilen özdeğer 5,06'dır. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,04 ile 5,11 arasında değişmektedir. ÇA ve SO yöntemleri için hesaplan özdeğer tam veri seti için hesaplanan özdeğerle aynıdır. En yüksek ve tam veri setinden elde edilen öz değere en uzak değer LBS (5,11) yönteminden elde edilirken, en düşük özdeğeri DDK (5,04) yöntemi vermiştir.

500 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4.14.'te görüldüğü gibidir.

**Tablo 4.14.** 500 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,75	,75	,73	,73	,73	,74	,75	,75	,75	,75
2	,79	,78	,78	,78	,78	,79	,78	,79	,79	,79
3	,74	,73	,73	,72	,73	,73	,73	,74	,74	,74
4	,72	,72	,72	,72	,72	,72	,72	,72	,72	,72
5	,78	,78	,78	,78	,78	,78	,78	,78	,78	,78
6	,77	,78	,77	,77	,77	,77	,78	,78	,77	,77
7	,72	,71	,71	,71	,70	,71	,71	,72	,72	,72
8	,74	,74	,73	,73	,73	,74	,74	,74	,74	,74
9	,74	,74	,73	,73	,73	,73	,74	,74	,74	,74
Açıklanan varyans	56,17	55,88	55,08	55,07	54,93	55,53	55,88	56,31	56,11	56,17
Özdeğer	5,06	5,03	4,96	4,96	4,94	5,00	5,03	5,07	5,06	5,06

Tablo 4.14'e göre, 500 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,72 ile ,79 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde RA ve ÇA yöntemleriyle elde edilen madde faktör yükleri tam veri seti ile aynı sonucu vermiştir. Bu yöntemlerden sonra tam veri setine en yakın madde faktör yükleri BM yöntemi ile elde edilmiştir. BM yönteminde 6.madde hariç tam veri seti ile aynı sonuçlar elde edilmiştir. SO, NDE ve LBS yöntemlerinde 4 madde hariç tam veri seti ile aynı sonuçlar elde edilirken, YNO, YNM ve DDK yöntemlerinde ise sadece 4, 5 ve 6.maddeler tam veri seti ile aynı madde faktör yük değerini vermiştir.

Tablo 4.14'e göre, 500 kişilik tam veri seti için elde edilen açıklanan varyans değeri 56,17'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri BM (56,31) yönteminden elde edilirken, en düşük değer DDK (54,93)

yönteminden elde edilmiştir. ÇA yöntemi için hesaplanan açıklanan varyans değeri ise tam veri seti ile aynıdır.

Tablo 4.14'e göre, 500 kişilik tam veri seti için elde edilen özdeğer 5,06'dır. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 4,94 ile 5,07 arasında değişmektedir. RA ve ÇA yöntemleri için hesaplan özdeğer tam veri seti için hesaplanan özdeğerle aynıdır. En düşük ve tam veri setinden elde edilen özdeğere en uzak değer DDK (4,94) yönteminden elde edilirken, en yüksek özdeğeri BM (5,07) yöntemi vermiştir.

500 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açılımlı faktör analizi sonuçları Tablo 4.15.'te görüldüğü gibidir.

**Tablo 4.15.** 500 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açılımlı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,75	,74	,74	,74	,74	,74	,76	,75	,75	,75
2	,79	,78	,78	,78	,78	,78	,78	,79	,79	,79
3	,74	,72	,72	,72	,72	,72	,75	,73	,73	,73
4	,72	,72	,71	,71	,70	,72	,73	,73	,72	,72
5	,78	,77	,77	,77	,77	,77	,77	,78	,77	,78
6	,77	,76	,77	,77	,76	,76	,76	,77	,77	,77
7	,72	,71	,71	,71	,71	,71	,72	,72	,72	,72
8	,74	,74	,74	,74	,73	,74	,75	,75	,74	,74
9	,74	,74	,74	,73	,73	,73	,75	,74	,74	,74
Açıklanan varyans	56,17	55,10	54,78	54,77	54,56	55,12	56,55	56,54	55,97	56,03
Özdeğer	5,06	4,96	4,93	4,93	4,91	4,96	5,09	5,09	5,04	5,04

Tablo 4.15'e göre, 500 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,72 ile ,79 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde tam veri setine en yakın sonucu ÇA yönteminin verdiği görülmektedir. Bu yöntemde 3.madde hariç tam veri seti ile aynı madde faktör yük değerleri elde edilmiştir. RA yönteminde 2 (3.ve 5.madde), BM yönteminde ise 3 (3,4 ve 8.madde) madde hariç tam veri seti ile aynı sonuçlar elde edilmiştir. SO (4,8 ve 9.madde) ve YNO (6,8 ve 9.madde) yöntemlerinde 3 madde, YNM (6.ve 8.madde) ve NDE (4.ve 8.madde) yöntemlerinde 2 madde, LBS (7.madde) yönteminde ise 1 madde tam veri seti ile aynı çıkarken; DDK yönteminde hiçbir madde tam veri seti ile aynı madde faktör yük değerini vermemiştir.

Tablo 4.15'e göre, 500 kişilik tam veri seti için elde edilen açıklanan varyans değeri 56,17'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri LBS (56,55) yönteminden elde edilirken, en düşük değer DDK (54,56) yönteminden elde edilmiştir. Tam veri setinden elde edilen değere en yakın hesaplanan değer ise ÇA (56,03) yönteminden elde edilmiştir.

Tablo 4.15'e göre, 500 kişilik tam veri seti için elde edilen özdeğer 5,06'dır. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 4,91 ile 5,09 arasında değişmektedir. RA ve ÇA (5,04) yöntemleri için hesaplan özdeğerler tam veri seti için hesaplanan özdeğere yakın çıkmıştır. En düşük ve tam veri setinden elde edilen özdeğere en uzak değer DDK (4,91) yönteminden elde edilirken, en yüksek özdeğeri LBS ve BM (5,04) yöntemleri vermiştir.

1000 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4.16.'da görüldüğü gibidir.

**Tablo 4.16.** 1000 kişilik tam veri setinden ve %5 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,75	,75	,75	,75	,75	,75	,76	,75	,75	,75
2	,80	,80	,80	,80	,80	,80	,80	,80	,80	,80
3	,77	,77	,77	,77	,77	,77	,78	,77	,77	,77
4	,76	,75	,76	,76	,76	,75	,76	,76	,76	,76
5	,77	,77	,77	,77	,77	,77	,77	,77	,77	,77
6	,80	,80	,80	,80	,80	,80	,80	,80	,80	,80
7	,77	,77	,77	,77	,77	,77	,77	,77	,77	,77
8	,72	,72	,72	,72	,72	,72	,72	,72	,72	,72
9	,72	,72	,72	,72	,72	,72	,72	,72	,72	,72
Açıklanan varyans	58,20	58,00	58,02	58,02	58,01	58,00	58,52	58,31	58,18	58,21
Özdeğer	5,24	5,22	5,22	5,23	5,22	5,22	5,27	5,25	5,24	5,24

Tablo 4.16’da görüldüğü üzere, 1000 kişilik tam veri seti için hesaplanan madde faktör yük değerleri ,72 ile ,80 arasında değişmekte olup bu veri seti için hesaplanan açıklanan varyans değeri 58,20 ve özdeğer ise 5,24’tür. Farklı kayıp veri yöntemlerinden elde edilen değerler tam veri setinden elde edilen bu değerler dikkate alınarak karşılaştırılmıştır.

Tablo 4.16’ya göre, 1000 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,72 ile ,80 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde SO ve NDE yöntemlerinde 4. madde, LBS yönteminde ise 1.ve 3.madde dışında, madde faktör yükleri tam veri seti ile aynı çıkmıştır.

Tablo 4.16’ya göre, 1000 kişilik tam veri seti için elde edilen açıklanan varyans değeri 58,20’dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan

varyans değeri LBS (58,52) yönteminden elde edilirken, en düşük değer SO ve NDE (58,00) yöntemlerinden elde edilmiştir. Tam veri setinden elde edilen değere en yakın hesaplanan değer ise ÇA (58,21) yönteminden elde edilmiştir.

Tablo 4.16'ya göre, 1000 kişilik tam veri seti için elde edilen özdeğer 5,24'tür. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,22 ile 5,27 arasında değişmektedir. RA ve ÇA yöntemleri için hesaplan özdeğer tam veri seti için hesaplanan özdeğerle aynıdır. En yüksek ve tam veri setinden elde edilen özdeğere en uzak değer LBS (5,27) yönteminden elde edilirken, en düşük özdeğeri SO, YNO, DDK ve NDE (5,22) yöntemleri vermiştir.

1000 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açılıyıcı faktör analizi sonuçları Tablo 4.17.'de görüldüğü gibidir.

**Tablo 4.17.** 1000 kişilik tam veri setinden ve %10 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açılıyıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Bekleniti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,75	,75	,75	,74	,75	,75	,75	,76	,75	,75
2	,80	,80	,80	,80	,80	,80	,80	,81	,81	,80
3	,77	,77	,77	,77	,77	,77	,78	,78	,78	,78
4	,76	,75	,75	,75	,75	,75	,76	,76	,76	,75
5	,77	,77	,77	,77	,76	,77	,77	,77	,77	,77
6	,80	,80	,80	,80	,80	,80	,81	,80	,80	,80
7	,77	,77	,76	,76	,77	,77	,77	,77	,77	,77
8	,72	,72	,72	,72	,72	,72	,72	,72	,72	,72
9	,72	,72	,72	,72	,72	,72	,72	,73	,73	,73
Açıklanan varyans	58,20	57,93	57,67	57,68	57,48	57,94	58,65	58,64	58,52	58,42
Özdeğer	5,24	5,21	5,19	5,19	5,17	5,22	5,28	5,28	5,27	5,26

Tablo 4.17'ye göre, 1000 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,72 ile ,80 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde SO ve NDE yöntemleri için 4.madde hariç tam veri seti ile aynı sonuçlar elde edildiği görülmektedir. YNO (4 ve 7.madde), DDK (4.ve 5.madde) ve LBS (3.ve 6.madde) yöntemlerinde 2 madde; YNM (1,4 ve 7.madde), RA (2,3 ve 9.madde), ÇA (3,4 ve 9.madde) yöntemlerinde 3 madde; BM (1,2,3 ve 9.madde) yönteminde ise 4 madde hariç tam veri seti ile aynı madde faktör yük değerleri elde edilmiştir.

Tablo 4.17'ye göre, 1000 kişilik tam veri seti için elde edilen açıklanan varyans değeri 58,20'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri LBS (58,65) yönteminden elde edilirken, en düşük değer DDK (57,48) yönteminden elde edilmiştir. Tam veri setinden elde edilen değere en yakın değer ise ÇA (58,42) yönteminden elde edilmiştir.

Tablo 4.17'ye göre, 1000 kişilik tam veri seti için elde edilen özdeğer 5,24'tür. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,17 ile 5,28 arasında değişmektedir. NDE (5,22) yöntemi için hesaplan özdeğer tam veri setine en yakın değeri vermiştir. En düşük ve tam veri setinden elde edilen özdeğere en uzak değer DDK (5,17) yönteminden elde edilirken, en yüksek özdeğeri LBS ve BM (5,28) yöntemleri vermiştir.

1000 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları Tablo 4.18.'de görüldüğü gibidir.

**Tablo 4.18.** 1000 kişilik tam veri setinden ve %20 oranında kayıp veri içeren veri seti için farklı kayıp veri yöntemlerinden elde edilen açımlayıcı faktör analizi sonuçları

Maddeler	Tam veri	Seriler ortalaması	Yakın noktaların ortalaması	Yakın noktaların medyanı	Doğrusal değer kestirimi	Noktanın doğrusal eğimi	Liste bazında silme	Beklenti maksimizasyonu	Regresyon ataması	Çoklu atama
1	,75	,74	,75	,75	,75	,74	,76	,75	,75	,75
2	,80	,80	,80	,80	,80	,80	,81	,81	,81	,81
3	,77	,77	,77	,77	,76	,77	,78	,78	,77	,78
4	,76	,75	,75	,75	,75	,75	,77	,76	,75	,76
5	,77	,76	,76	,75	,75	,76	,78	,77	,77	,77
6	,80	,79	,79	,79	,79	,79	,81	,80	,80	,80
7	,77	,76	,76	,76	,76	,76	,77	,77	,76	,76
8	,72	,72	,72	,72	,72	,72	,72	,72	,72	,72
9	,72	,72	,72	,72	,72	,72	,74	,73	,72	,72
Açıklanan varyans	58,20	57,32	57,04	56,92	56,86	57,33	59,32	58,71	58,12	58,20
Özdeğer	5,24	5,16	5,13	5,12	5,12	5,16	5,34	5,28	5,23	5,24

Tablo 4.18'e göre, 1000 kişilik tam veri seti için elde edilen madde faktör yük değerleri ,72 ile ,80 arasında değişmektedir. Farklı kayıp veri yöntemleri çerçevesinde elde edilen madde faktör yük değerleri incelendiğinde tam veri setine en yakın sonuçların BM, RA ve ÇA yöntemlerinden elde edildiği görülmektedir. BM (2,3 ve 9. madde), RA (2,4 ve 7.madde) ve ÇA (2,3 ve 7.madde) yöntemleri için 3'er madde hariç tam veri seti ile aynı madde faktör yük değerleri elde edilmiştir. YNM ve YNO yöntemlerinde 4,5,6 ve 7.maddeler dışında kalan maddeler tam veri setiyle aynı sonucu vermiştir. SO (1,4,5,6 ve 7.madde), DDK (3,4,5,6 ve 7.madde) ve NDE (1,4,5,6 ve 7. madde) yöntemlerinde ise 5 madde tam veri setinden farklılaşmaktadır. Tam veri seti ile en az benzerlik gösteren yöntem ise LBS yöntemi olup sadece 2 madde tam veri seti ile benzerlik göstermektedir.

Tablo 4.18'e göre, 1000 kişilik tam veri seti için elde edilen açıklanan varyans değeri 58,20'dir. Farklı kayıp veri yöntemleri için elde edilen en büyük açıklanan varyans değeri LBS (59,32) yönteminden elde edilirken, en düşük değer DDK (56,86) yönteminden elde edilmiştir. ÇA (58,20) yöntemi ise tam veri setinden elde edilen değer ile aynı sonucu vermiştir.

Tablo 4.18'e göre, 1000 kişilik tam veri seti için elde edilen özdeğer 5,24'tür. Farklı kayıp veri yöntemleri için elde edilen özdeğerler 5,12 ile 5,34 arasında değişmektedir. ÇA yöntemi için hesaplanan özdeğer tam veri seti için hesaplanan özdeğerle aynıdır. En düşük ve tam veri setinden elde edilen özdeğere en uzak değer YNM ve DDK (5,12) yöntemlerinden elde edilirken, en yüksek özdeğeri LBS (5,34) yöntemi vermiştir.

## BÖLÜM V

### 5.Sonuç ve Öneriler

Bu bölümde araştırmanın bulguları doğrultusunda elde edilen sonuç ve önerilere yer verilmiştir.

#### 5.1. Sonuç

Bu araştırmada normal dağılıma sahip tek faktörlü tam veri setleri ile değişen oranlarda (%5, %10, %20) kayıp veri içeren çeşitli büyüklükteki (n=200, n=500, n=1000) veri setleri için farklı kayıp veri yöntemlerinden elde edilen yeni veri setleri betimsel istatistik, güvenilirlik ve geçerlik açısından karşılaştırarak, hangi koşullarda hangi kayıp veri yönteminin kullanılması gerektiğine ilişkin çıkarımlarda bulunmak amaçlanmıştır. Araştırmada dokuz farklı kayıp yöntemi ele alınmıştır. Araştırmada kullanılan veriler normal dağılıma ve tek faktörlü bir yapıya sahip gerçek verilerdir. Tam veri seti ile kayıp veri yöntemlerinden elde edilen yeni veri setlerinden elde edilen değerlerin karşılaştırılmasında TROK mekanizması temel alınmıştır.

Araştırmada ulaşılan sonuçlar, her bir alt problem için ayrı ayrı olmak üzere aşağıda verilmiştir.

Birinci alt probleme ilişkin bulgulara dayalı olarak ulaşılan sonuçlar;

Farklı büyüklük ve farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin kullanılması neticesinde elde edilen ortalama değerleri ile tam veri setlerinden elde edilen ortalama değerleri karşılaştırıldığında, genel olarak LBS yönteminden elde edilen

değerler tam veri setlerinden elde edilenlere göre daha yüksektir. Diğer kayıp veri yöntemlerinden elde edilen değerler de tam veri setlerinden elde edilen değerlere göre değişkenlik göstermektedir. Tam veri setlerinden edilen ortalama değerlerine yakın değerler genel olarak ÇA yönteminden elde edilmiştir. Her ne kadar tam veri setlerinden ve kayıp veri yöntemlerinden elde edilen ortalama değerleri betimsel anlamda farklılıklar gösterse de yapılan t testi analizi ortalamalar arasındaki farklılığın manidar olmadığını göstermektedir. Sonuç olarak, farklı kayıp veri yöntemlerinin tamamının, farklı büyüklük ve farklı kayıp veri oranlarının tamamı için, tam veri setlerinden elde edilen ortalama değerlerle benzer sonuçlar verdiği görülmüştür.

Farklı büyüklük ve farklı kayıp veri oranları için çeşitli kayıp veri yöntemlerinin kullanılması neticesinde elde edilen standart sapma değerleri ile tam veri setlerinden elde edilen standart sapma değerleri karşılaştırıldığında, LBS yönteminin tam veri setlerinden elde edilenlere göre daha yüksek, yaklaşık değer atama yöntemlerinin ise daha düşük değerler verdiği görülmüştür. Böyle olmakla birlikte, farklı kayıp veri yöntemlerinin her birinden elde edilen değerlerle tam veri setlerinden elde edilen değerlerin birbirlerine büyük ölçüde benzedikleri ve aralarındaki farklılıkların oldukça küçük olduğu görülmüştür. Sonuç olarak, farklı kayıp veri yöntemlerinin tamamının, farklı büyüklük ve farklı kayıp veri oranlarının tamamı için, tam veri setlerinden elde edilen standart sapma değerleriyle benzer sonuçlar verdiği görülmüştür.

Çarpıklık değerleri açısından, 200 kişilik veri setinde tüm farklı kayıp veri oranları için BM, RA ve ÇA yöntemleri tam veri setinden elde edilen değere en yakın değerleri verirken, en yüksek değerlerin farklı kayıp veri oranlarına göre LBS ve DDK yöntemleri arasında değiştiği görülmüştür. 500 kişilik tam veri setinden elde edilen çarpıklık değerlerine göre en yüksek değerler %10 kayıp veri oranı için RA, diğer kayıp veri oranları için LBS yönteminden elde edilirken, kayıp veri oranlarına göre tam veri setinden elde edilen değere en yakın değeri veren kayıp veri yöntemleri BM ve ÇA yöntemleri olmuştur. 1000 kişilik veri setinde %5 kayıp veri oranı için DDK, % 10 kayıp veri oranı için SO ve YNO, %20 kayıp veri oranı için ise LBS yöntemleri tam veri setinden elde edilen değere göre yüksek değerler vermiştir. Tam veri setinden elde edilen değere göre en düşük değerler %5 kayıp veri oranı için RA, %20 kayıp veri oranı

için DDK yöntemlerinde görülmüştür. Tam veri setlerinden elde edilen değerlere en yakın değerleri veren kayıp veri yöntemleri kayıp veri oranlarına göre LBS, RA ve ÇA yöntemleri olmuştur. Ortalama ve standart sapma değerlerinde olduğu gibi, yapılan karşılaştırmalar, farklı kayıp veri yöntemlerinin her birinden elde edilen çarpıklık değerleri ile tam veri setlerinden elde edilen değerlerin birbirlerine büyük ölçüde benzediklerini ve aralarındaki farklılıkların oldukça küçük olduğunu göstermiştir. Sonuç olarak, farklı kayıp veri yöntemlerinin tamamının, farklı büyüklük ve farklı kayıp veri oranlarının tamamı için, tam veri setlerinden elde edilen çarpıklık değerleriyle benzer sonuçlar verdiği görülmüştür.

Basıklık değerleri bakımından, 200 kişilik veri seti için, tam veri setinden elde edilen değere göre en yüksek basıklık değerlerini, farklı kayıp veri oranlarına göre, LBS ve DDK yöntemleri vermiştir. 500 ve 1000 kişilik veri setleri için, tam veri setlerinden elde edilen değerlere göre en yüksek basıklık değerleri, farklı kayıp veri oranları için, LBS ve yaklaşık değer atama yöntemleri arasında değişmektedir. Farklı büyüklükteki veri setleri için, tam veri setlerinden elde edilen değerlere göre genel olarak en düşük basıklık değerleri farklı kayıp veri oranları için LBS, BM, RA ve ÇA yöntemleri arasında değişse de BM, RA ve ÇA yöntemleri tam veri setinden elde edilenlere en yakın değerleri veren yöntemler olmuşlardır. Yapılan karşılaştırmalar farklı kayıp veri yöntemlerinin her birinden elde edilen basıklık değerleri ile tam veri setlerinden elde edilen değerlerin birbirlerine büyük ölçüde benzediklerini ve aralarındaki farklılıkların oldukça küçük olduğunu göstermiştir. Sonuç olarak, farklı kayıp veri yöntemlerinin tamamının, farklı büyüklük ve farklı kayıp veri oranlarının tamamı için, tam veri setlerinden elde edilen basıklık değerleriyle benzer sonuçlar verdiği görülmüştür.

Araştırmada ele alınan betimsel istatistik değerlerine ilişkin yapılan karşılaştırmalar bir bütün olarak ele alındığında, genel olarak tüm farklı örneklem büyüklüğü ve kayıp veri oranları için benzer sonuçlar elde edilmiştir. LBS yöntemi için hesaplanan değerler tam veri setinden elde edilen değerlerden en fazla farklılık gösteren değerler olurken, ÇA yöntemi genelde tam veri setinden elde edilenlere en yakın değerleri veren yöntem olmuştur. BM ve RA yöntemleri de ÇA yönteminden sonra tam veri setinden elde edilen değerlere en yakın değerleri veren diğer yöntemler olarak ön

plana çıkmaktadırlar. Yaklaşık değer atama yöntemleri genellikle tam veri setinden elde edilen değerlere göre düşük değerler vermişlerdir. Farklı kayıp veri yöntemlerinden elde edilen değerler arasında tam veri setlerinden elde edilenlere göre farklılıklar gözlemlense de bu farklılıkların çok küçük değerler biçiminde olduğu görülmüştür. Örneğin, ortalamalar arasındaki farklılıkların manidarlığına ilişkin yapılan t testi analizi manidar farklılığın sadece 1000 kişilik veri seti için %10 kayıp veri durumunda BM ve RA yöntemlerinde olduğu görülmüştür. Bu durum Bal (2003)'ün çalışmasında da görülmektedir. Yöntemler arasında büyük farklılıklar bulunmasa da ÇA yöntemi genellikle tam veri setlerinden elde edilen değerlere daha yakın sonuçlar verdiği için, araştırmada dikkate alınan koşullar altında, diğer yöntemlere göre en kullanılabilir yöntem olarak ön plana çıkarırken, LBS yönteminin diğer yöntemlere göre en az kullanılabilir yöntem olduğu söylenebilir.

İkinci alt probleme ilişkin bulgulara dayalı olarak ulaşılan sonuçlar;

Farklı örneklem büyüklüğü ve farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin kullanılması neticesinde elde edilen Cronbach alpha güvenilirlik katsayıları ile tam veri setlerinden elde edilen katsayıların genel olarak farklılaşmakla birlikte bu farklılıkların çok küçük değerler şeklinde olduğu görülmüştür. Tüm farklı örneklem büyüklüklerinde, kayıp veri oranının düşük olduğu durumda (%5) tam veri setlerinden elde edilen değerlere çok yakın güvenilirlik katsayıları elde edilirken, kayıp veri miktarı arttıkça elde edilen katsayılar tam veri setlerinden elde edilen değerlerden az da olsa farklılaşmıştır. Özellikle kayıp veri oranının yüksek olduğu durumlarda yaklaşık değer atama yöntemlerinden elde edilen güvenilirlik katsayıları, tam veri setinden elde edilenlerden daha düşük çıkmıştır. LBS yöntemi kayıp veri oranının düşük olduğu durumda (%5) genelde tam veri setlerinden elde edilen katsayılara yakın veya aynı katsayı değerini verirken, kayıp veri oranı arttıkça tam veri setlerinden elde edilen katsayılardan daha yüksek değerler almıştır. Bu bulgu Demir (2013)'ün çalışmasıyla da tutarlıdır. Yaklaşık değer atama yöntemlerinde kayıp veri oranının fazla olduğu durumlarda yapılan kestirimlerin tam veri setinden elde edilen değerlere göre daha düşük olduğu görülmektedir ve bu bulgu Çokluk ve Kayri (2011) ve Demir (2013)'ün yaklaşık değer atama yöntemlerinden bazılarının kullanıldığı çalışması ile tutarlıdır.

BM, RA ve özellikle de ÇA yöntemleri tüm farklı örneklem büyüklüğü ve kayıp veri oranları için tam veri setlerinden elde edilen değerlere çok yakın veya aynı değerleri vermiştir. BM yönteminin tam verilerden elde edilen güvenilirlik katsayılarına benzer kestirimlerde bulunabildiğine ilişkin elde edilen bu bulgu Enders (2004)'ün çalışmasını destekler niteliktedir. Bunun yanı sıra, kayıp veri oranının yüksek olduğu durumda, ÇA yönteminin diğer kayıp veri yöntemlerine göre daha iyi sonuçlar vermesi Granberg-Rademacker (2007), Leite ve Beretvas (2010) ve Young ve diğerleri (2011)'in çalışmasıyla da tutarlık göstermektedir. Cronbach alpha güvenilirlik katsayıları açısından tam veri setlerinden elde edilen değerlerle kayıp veri yöntemlerinden elde edilen değerler arasında belli düzeyde farklılıklar gözlemlense de Fisher z testi bu farklılıkların manidar olmadığını göstermiştir. Bu durum kayıp veri yöntemlerinin, araştırmada dikkate alınan koşullar altında, aynı kullanılabilirlik düzeyinde olduğunu göstermesi bakımından önemli görülebilir. Ancak, yine de, özellikle kayıp veri oranının yüksek olduğu durumlarda yaklaşık değer atama yöntemleri ile elde edilen katsayıların tam veri setlerinden elde edilen değerlerden daha düşük olmasından dolayı, kayıp verinin fazlalığı durumunda, bu yöntemlerin olumsuz etki yaratabileceği düşünülmektedir.

Üçüncü alt probleme ilişkin bulgulara dayalı olarak ulaşılan sonuçlar;

Açımlayıcı faktör analizi bağlamında yapılan karşılaştırmalar farklı örneklem büyüklüğü ve farklı kayıp veri oranları için tüm kayıp veri yöntemlerinin, tam veri setlerine benzer biçimde, tek faktörlü yapıyı gösterdiği görülmüştür. Bu bulgu Çokluk ve Kayri (2011)'in yaklaşık değer atama yöntemlerine yer verdiği çalışması ve Chen ve diğerleri (2012)'nin SO, NDE, DDK, LBS, BM ve RA yöntemlerine yer verdiği çalışması ile de tutarlık göstermektedir.

Farklı büyüklükteki veri setlerinin tamamı için kayıp veri oranı arttıkça madde faktör yük değerlerinin tam veri setlerinden elde edilen değerlere benzerliği küçük bir miktar azalsa da düşük ve yüksek faktör yük değeri veren maddelerin neredeyse tüm koşullar altında benzerlik gösterdiği gözlemlenmiştir. BM, RA ve ÇA yöntemleri farklı kayıp veri oranlarının tamamı için, özellikle de kayıp veri oranının yüksek olduğu durumlar için, genel olarak tam veri setlerinden elde edilen değerlere en yakın madde

faktör yük değerleri vermişlerdir. Yaklaşık değer atama ve silme yöntemlerinde ise farklı büyüklükteki veri setlerinin tamamı için kayıp veri oranı arttıkça, madde faktör yük değerlerinin tam veri seti ile benzerliği de azalmıştır.

Açıklanan varyans değerleri ve özdeğerler bakımından, genel olarak, yaklaşık değer atama yöntemlerinden elde edilen değerler tam veri setlerinden elde edilen değerlerden az da olsa düşük bulunmuştur. Bu bulgu Çokluk ve Kayri (2011)'in çalışması ile benzerlik göstermektedir. Kayıp veri yöntemleri içinde özellikle DDK yöntemi tam veri setlerinden elde edilen değerlere göre en düşük değerleri veren yöntem olmuştur. LBS yöntemi genel olarak en yüksek değerleri veren yöntem olsa da bazı durumlarda BM ve RA yöntemlerinin de en yüksek değerleri verdiği görülmüştür. BM ve RA yöntemleri ÇA yönteminden sonra tam veri setlerinden elde edilen değerlere en yakın değerleri veren yöntemlerdir. Yani açıklanan varyans ve özdeğerler bakımından da BM, RA ve ÇA yöntemleri tam veri setlerinden elde edilen değerlere benzer değerler verebilmeleri bakımından ön plana çıkmışlardır.

Kayıp veri yöntemleri genel olarak değerlendirildiğinde, LBS yöntemi genelde tam veri setlerinden elde edilen değerlerden en fazla farklılaşan değerleri veren yöntem olmuştur. Bu bulgu yapılan bazı çalışmalarla da desteklenmektedir (Acock, 2005; Baraldi ve Enders, 2010; Ginkel ve diğerleri, 2010; Hawthorne ve Elliot, 2005; Peng ve diğerleri, 2006; Roth, 1994; Roth ve Switzer, 1995). LBS yöntemi örneklem sayısının düşmesine neden olduğu için tip II hata oluşumu riskini arttırmakta ve bundan dolayı istatistiksel olarak önemli sonuçlara ulaşmanın ihtimalini azaltmaktadır (Buhi ve diğerleri, 2008; Streiner, 2002). Bunun yanı sıra yanlış parametre tahminlerine neden olmaktadır (Buhi ve diğerleri, 2008; Ginkel ve diğerleri, 2010; Graham, 2009; Imtiaz, 2007; Peng ve diğerleri, 2006). Bu durumla özellikle veri yapısının TROK olmadığı durumlarda karşılaşılmaktadır (Brand, 1999; Dragset, 2009; Dural, 2010; Haitovsky, 1968; Imtiaz, 2007; King, Honaker, Joseph ve Scheve, 2001; Leite ve Beretvas, 2010; Peng ve diğerleri, 2006). Ancak veri yapısının TROK olduğu durumlarda LBS yöntemini uygulamak için kayıp veri içeren satırların silinmesiyle elde edilen yeni veri seti, rastgele örneklem seçimi ile elde edilmiş sayılacağından, bu yöntemin yanlış sonuçlar vermeyeceği (Ginkel ve diğerleri, 2010; Schlomer ve diğerleri, 2010) ve tek

olumsuz yanının istatistiksel güç kaybı ve örneklem sayısında azalma olduğu ifade edilmiştir (Acock, 2005; Pokropek, 2011; Saunders vd., 2006; Schlomer ve diğerleri, 2010). Bu yöntem, bazı kaynaklarda kayıp veri sorunu ile başetmede kullanılabilir en zayıf yöntem olarak tanımlansa da (Wilkinson ve The APA Task Force on Statistical Inference, 1999) veri yapısının TROK olduğu durumda geçerli kestirimler yaptığından dolayı iyi bir yöntem olarak da tanımlanmaktadır. Hatta regresyon analizi gibi bazı durumlarda çoklu atama ve en çok olabilirlik yöntemlerine göre daha iyi sonuçlar verdiği belirtilerek (Allison, 2002, s7) kullanılması gerektiğine ilişkin önerilerde bulunulmuştur (Roth ve Switzer, 1995). Bu çalışma kapsamında diğer yöntemlere göre tam veri setlerinden elde edilen değerlerden en fazla farklılık gösteren değerleri verse de kayıp veri yapısının TROK olmasından dolayı bu farklılıkların çok da büyük olmadığı görülmüştür.

Çalışma genelinde yaklaşık değer atama yöntemlerine bakıldığında, kayıp veri oranının düşük olduğu durumlarda, bu yöntemlerin genel olarak tam veri setlerinden elde edilen değerlere yakın veya aynı değerleri verdiği; kayıp veri oranı arttıkça da farklılaşan değerler verdiği görülmüştür. Bu bulgu, yapılan bazı araştırmalarla benzerlik göstermektedir (Chen ve diğerleri, 2012; Çokluk ve Kayrı, 2011).

Araştırmada ele alınan 9 farklı kayıp veri yönteminden, farklı örneklem büyüklüğü ve farklı kayıp veri oranlarının tamamı için, tam veri setlerinde elde edilen değerlere en yakın veya aynı değerleri büyük ölçüde BM, RA ve ÇA yöntemleri vermiştir. Bu çalışmada özellikle kayıp veri oranının yüksek olduğu durumlarda bu üç yöntemin, diğer yöntemlere göre tam veri setlerinden elde edilen değerlere en yakın sonuçlar verdiği görülmüştür.

RA yöntemi, bazı durumlarda tam veri setlerinden elde edilen değerlerden farklı değerler verse de, bu değerler genel olarak tam veri setlerinden elde edilen değerlerle ya çok benzer ya da aynıdır. Bu bulgu Suraphee, Raksmanee, Busaba, Chaisorn, ve Nakornthai (2006) ve Roth ve Switzer (1995)'in çalışmasıyla da desteklenmektedir. Bu çalışma sonuçlarıyla da tutarlı olarak, hata teriminin eklendiği regresyon ataması yönteminin silme yöntemleri (Hertel, 1976; Kim ve Curry, 1977; McDonald, Thurston

ve Nelson, 2000; Raymond ve Roberts, 1987; Roth ve Switzer, 1995; Roth, Switzer III ve Switzer, 1999; Roth, Campion ve Jones, 1996) ve bazı basit atama yöntemlerine göre daha iyi sonuçlar verdiği alanyazında yapılan çalışmalarda belirtilmiştir (McDonald ve diğerleri, 2000; Musil, Warner, Yobas ve Jones, 2002; Roth ve Switzer, 1995; Roth ve diğerleri, 1999; Schafer ve Graham, 2002; Streiner, 2002). Hatta veriler arasında yüksek korelasyon varsa yöntemin daha iyi sonuçlar verdiği belirtilmiştir (Durrant; 2005; Jing, 2012; McDonald ve diğerleri, 2000; Sezgin ve Çelik, 2013). Ancak, RA yönteminin en önemli dezavantajlarından biri veriler arasında olan ilişkiyi daha da kuvvetlendirdiği için hatalı çıkarımlar yapılmasına neden olmasıdır (Brown ve Kros, 2003; Newman, 2003).

BM yöntemi de RA yöntemi gibi araştırmacılar tarafından sık kullanılan bir kayıp veri yöntemidir. Bu çalışmada, BM yöntemi, silme ve yaklaşık değer atama yöntemlerine göre, tam veri setlerinden elde edilen değerlere yakın veya tam veri setleri ile aynı değerleri vermiştir. Bu yöntemin bazı basit atama yöntemleri (Bernaards ve Sijtsma, 2000; Graham ve diğerleri, 1994) ve liste bazında silme yöntemine göre daha iyi sonuçlar ürettiği de yapılan çalışmalarla desteklenmektedir (Bal, 2003; Bernaards ve Sijtsma, 2000; Cheema, 2012; Enders, 2004; Musil ve diğerleri, 2002; Peng ve diğerleri, 2006; Streiner, 2002). Alanyazında ortalama, standart sapma, güvenilirlik analizi ve açılımlayıcı faktör analizi gibi durumlarda BM yönteminin hem uygulamasının kolay hem de iyi sonuçlar verdiği belirtilmiştir (Graham, 2009). Veri yapısının TROK veya ROK olduğu durumda yansız parametre değerleri ürettiği de belirtilmektedir (Baraldi ve Enders, 2010; Dragset, 2009). BM yönteminin tam veri setlerinden elde edilen değerlere yakın değerler vermesinin nedeni, en çok olabilirlik yöntemi ile veri setindeki tüm bilgileri kullanmasıdır (Graham, 2009).

ÇA yöntemi bazı farklılıklar gözlemlense de tam veri setlerinden elde edilen değerlere en yakın değerleri veren yöntem olmuştur. ÇA yöntemi ile yapılan birçok çalışmada yöntem, farklı kayıp veri yöntemleri ile karşılaştırılmıştır. Bu çalışmalarda, ÇA yönteminin diğer yöntemlere göre daha iyi tahmin değerleri ürettiği (Arnold ve Kronmal, 2002, Baraldi ve Enders, 2010; Barzi ve Woodward, 2004; Buhi ve diğerleri, 2008; Durrant, 2005; Graham ve diğerleri, 1994; Jing, 2012; Leite ve Bretvas, 2010;

Newman, 2003; Olinsky ve diğ erleri, 2003; Pigott, 2001; Pokropek, 2011; Streiner, 2002; Wayman, 2003) ve veri yapısının TROK ve ROK oldu ğ u durumlarda da yansız parametre tahminleri yaptı ğ ı belirtilmektedir (Baraldi ve Enders, 2010; Donders, Van der Heijden, Stijnen ve Moons, 2006). Ç A yöntemi bazı ek hesaplama iş lemleri gerektirdi ğ i için arařtırmacılar tarafından çok tercih edilen bir yöntem de ğ ildir. Ancak, geliřtirilen birç ok paket programla birlikte yöntemin uygulaması kolaylařmıřtır. Parametre tahminlerinde de en çok olabilirlik yöntemi ile en do ğ ru sonucu veren yöntem olmasından dolayı geleneksel yöntemlere göre kullanılması daha çok önerilmektedir (Baraldi ve Enders, 2010; Brand, 1999; Demir, 2013; Enders, 2013; Kang, 2013; Young, Weckmen ve Holland, 2011). Bunun yanı sıra Ç A yöntemi özellikle kayıp veri miktarının fazla oldu ğ u durumda tam veri setine yakın sonuçlar vermesinden dolayı, kayıp veri miktarının fazla oldu ğ u durumlarda di ğ er yöntemlere göre tercih edilebilir.

Kayıp veri yöntemlerinden herhangi birinin her kořul altında do ğ ru sonuç verdi ğ ine iliřkin bir bulguya rastlamak zordur. Kayıp veri yöntemlerinin etkilili ğ i karřılařılan kayıp veri mekanizmasına, kayıp veri miktarına, örnekleme büyüklü ğ une vb. etkenlere göre de ğ iřmektedir. Bu çalıřmada ele alınan kayıp veri yöntemleri TROK mekanizması temelinde de ğ erlendirilmiřtir. Veri yapısı TROK oldu ğ u zaman elde edilen de ğ erler tam veri setlerinden elde edilen de ğ erlerle aynı çıkmayabilir ancak yapılan kestirimler, kayıp veri yapısı ihmal edilebilir oldu ğ u için yansız olmaktadır (Bernaards ve Sijtsma, 2000; Buhi ve di ğ erleri, 2008; Graham, 2009). Bu çalıřmada kayıp veri yöntemleri ile oluřturulan yeni veri setleri ile tam veri setlerinden elde edilen de ğ erler arasında büyük benzerliklerin oluřu, ya da bir bařka deyiřle, çok küçük farklılıkların bulunuşunun nedeni veri yapısının TROK olmasıyla açıklanabilir. Sonuç olarak, veri yapısı TROK oldu ğ unda yöntemlerin birbirlerine göre üstünlü ğ ü tartıřmaya açık olsa da BM, RA ve Ç A yöntemlerinin di ğ er yöntemlere yüksek performans gösterdi ğ i açıktır.

## 5.2. Öneriler

Araştırmadan elde edilen bulgular doğrultusunda aşağıdaki öneriler dikkate alınabilir;

1. Kayıp veri miktarının fazla olduğu durumda çoklu atama yönteminin kullanılması diğer yöntemlere göre tercih edilmelidir.
2. Kayıp veri miktarının düşük olduğu durumda liste bazında silme yöntemi hariç diğer yöntemler tercih edilmelidir.
3. Silme ve yaklaşık değer atama yöntemlerine göre beklenti maksimizasyonu, regresyon ataması ve çoklu atama yöntemleri daha çok tercih edilmelidir.

Bu çalışma kapsamında farklı örneklem büyüklüğü ve kayıp veri oranına göre kayıp veri yöntemleri karşılaştırılmıştır. Çalışmada gerçek veriler kullanılmıştır ve çalışma kapsamında kullanılan veriler tek faktörlü ve normal dağılıma sahiptir. Kayıp veriler ise TROK mekanizmasındadır. Gelecekteki araştırmalara yönelik olarak çok faktörlü, çarpık dağılıma sahip, farklı örneklem büyüklüğü ve farklı kayıp veri oranlarına göre ROK veya IE mekanizmasına sahip kayıp veriler üzerinde farklı yöntemler kullanarak yöntemlerin etkinliği karşılaştırılabilir.

## KAYNAKÇA

- Acock, A.C. (2005). Working with missing values. *Journal of Marriage and Family*, 67, 1012-1028.
- Acuna, R. ve Rodriguez, C. (2004). The treatment of missing values and its effect in the classifier accuracy. Banks, D., House, L., McMorris, F.R., Arabie, P., Gaul, W. (Eds.). *Classification, clustering and data mining applications* (ss.639-649) içinde Germany: Springer.
- Agresti, A. ve Finlay, B. (1997). *Statistical methods for the social sciences (4.Edition)*. USA: Pearson Prentice Hall.
- Allison, P.D. (2002). *Missing data*. CA: Sage Publications.
- Allison, P.D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112 (4), 545-557, DOI: 10.1037/0021-843X.112.4.545.
- Allison, P.D. (2009). *Missing data* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 72-89). London: Sage Publication.
- Alosh, M. (2009). The impact of missing data in a generalized integer-valued autoregression model for count data. *Journal of Biopharmaceutical Statistics*, 19(6), 1039-1054, DOI: 10.1080/10543400903242787.
- Akbař, U. (2014). *Farklı örnekleme büyüklüğü ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi*. Yayımlanmamış Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- Akhun, İ. (1994). *İstatistiksel formüller ve tablolar (4.Baskı)*. Ankara: Hacettepe Üniversitesi, Eğitim Faültesi
- Andridge, R.R. ve Little, R.J.A. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78(1), 40–64, DOI:10.1111/j.1751 5823.2010.00103.x
- Anıl, D. (2011). Türkiye'nin Pisa 2006 fen bilimleri başarısını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 11(3), 1253-1266.
- Arnold, A.M. ve Kronmal, R.A. (2002). Multiple imputation of baseline data in the cardiovascular health study. *American Journal of Epidemiology*, 157(1), DOI: 10.1093/aje/kwf156.
- Bal, C. (2003). *Çok gruplu veri setlerinde eksik gözlem sorununun çözümlenmesi ve sağlık alanında bir uygulama*. Yayımlanmamış Doktora Tezi, Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.
- Baraldi, A.N. ve Enders, C.K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology*, 48, 5–37.
- Barzi, F. ve Woodward, M. (2004). Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology*, 160(1), 34-45, DOI: 10.1093/aje/kwh175.
- Baygül. A. (2007). *Kayıp veri analizinde sıklıkla kullanılan etkin yöntemlerin değerlendirilmesi*. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Üniversitesi, Sağlık Bilimleri Enstitüsü, İstanbul.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması (2. Baskı)*. Ankara: Pegem Akademi.

- Baykul, Y. ve Güzeller, C.O. (2013). *Sosyal bilimler için istatistik: SPSS uygulamalı*. Ankara: Pegem Akademi.
- Bennet, D.A. (2001). How can I deal with missing data in my study?. *Australian and New Zealand Journal of Public Health*, 25(5), 464-469
- Bernaards, C.A. ve 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, DOI: 10.1207/S15327906MBR3503\_03.
- Brand, J.P.L. (1999). *Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets*. Print Partners Ispkamp.
- Brown, M.L. ve Kros, J. F. (2003). Data mining and the impact of missing data. *Industrial Management & Data System*, 103 (8), 611- 621, DOI: 10.1108/02635570310497657.
- Buhi, E.R., Goodson, P. ve Neilands, T.B. (2008). Out of sight , not out of mind: Strategies for handling missing data. *American Journal of Health Behavior*, 32 (1), 83-92.
- Büyüköztürk, Ş. (2012). *Sosyal bilim için veri analizi el kitabı (17.Baskı)*. Ankara: Pegem Akademi.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, 32, 470-483.
- Carpita, M. ve Manisera, M. (2011). On the imputation of missing data in surveys with likert- type scales. *Journal of Classical*, 28, 93-112, DOI: 10.1007/s00357-011-9074-z

- Cheema, J. (2012). *Handling missing data in educational research using SPSS*. Unpublished doctoral dissertation, George Mason University, USA.
- Chen, S.F., Wang, S. ve Chen, Y.C. (2012). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications*, 39, 4026–4031.
- Cheung, M.W.L. (2007). Comparison of methods of handling missing time- invariant covariates in latent growth models unders the assumption of missing completely at random. *Organizational Research Methods*, 10(4), 609-634, DOI: 10.1177/1094428106295499.
- Collins, L.M., Schafer, J.L. ve Kam, C.M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330-351, DOI: 10.1037//1082-989X.6.4.330.
- Cool, A.L. (Ocak, 2000). *A review of methods for dealing with missing data*. Annual Meeting of the Southwest Educational Research Association'da sunulmuş bildiri. Dallas.
- Cumming, P. (2013). Missing data and multiple imputation. *Clinical Review & Education*, 167 (7), 656-661
- Çakmur, H. (2012). Araştırmalarda ölçme- güvenirlik- geçerlik. *TAF Preventive Medicine Bulletin*, 11(3), 339-344.
- Çokluk, Ö. ve Kayrı, M. (2011). Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenirliği üzerindeki etkisi. *Kuram ve Uygulamada Eğitim Bilimleri*, 11 (1), 289-309.
- Demir, E. ve Parlak, B. (2012). Türkiye'de eğitim araştırmalarında kayıp veri sorunu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3 (1), 230-241.

- Demir, E. (2013). Kayıp verilerin varlığında çoktan seçmeli testlerde madde ve test parametrelerinin kestirilmesi: SBS örneği. *Eğitim Bilimleri Araştırmaları Dergisi*, 3 (2), 47-68.
- Dempster, A.P., Laird, N.M. ve Rubin, D.B.(1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1-38.
- Donders, A.R.T., van der Heijden, G.J.M.G., Stijnen, T. ve Moons, K.G.M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59, 1087-1091, DOI: 10.1016/j.jclinepi.2006.01.014.
- Dragset, I.G.(2009). *Analysis of longitudinal data with missing values. Methods and Applications in Medical Statistics*. Yayınlanmamış Yüksek Lisans Tezi, Norwegian University of Science and Technology, Department of Mathematical Sciences.
- 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, DOI: 10.1080/10705519809540086
- Dural, S. (2010). *Farklı kayıp veri tekniklerinin çok göstergeli örtük gelişme modelleri üzerindeki etkisi*. Yayınlanmamış Doktora Tezi, Ege Üniversitesi, Sosyal Bilimler Enstitüsü, İzmir.
- Durrant, G.B.(2005). *Imputation methods for handling item-nonresponse in the social sciences: A methodological review*. NCRM(NCRM Working Paper Series, (NCRM-002)). University of Southampton.
- Downey, R.G. ve King, C.V. (1998). Missing Data in Likert Ratings: A Comparison of Replacement Methods. *The Journal of General Psychology*, 125 (2), 175-191, DOI: 10.1080/00221309809595542

- Enders, C.K. (2001(a)). The performance of full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61(5), 713-740, DOI: 10.1177/0013164401615001.
- Enders, C.K. (2001 (b)). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, 8(1), 128-141, DOI: 10.1207/S15328007SEM0801\_7.
- Enders, C.K. ve Bandalos, D.L. (2001). The relative performance of full information maksimum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8(3), 430-457.
- 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, DOI: 10.1177/0013164403261050.
- Enders, C.K. (2010). *Applied missing data analysis*. USA: The Guilford Press.
- Enders, C.K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives*, 7 (1), 27- 31.
- Ercan, İ. ve Kan, İ. (2004). Ölçeklerde güvenilirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30 (3), 211-216.
- Erkuş, A (2006). *Sınıf öğretmenleri için ölçme ve değerlendirme, kavramlar ve uygulamalar*. Ankara: Ekinoks.
- Field, A. (2009). *Discovering statistics using SPSS* (3. Edition). London: Sage Publication

- Finch, H., & Margraf, M. (2008). *Imputation of categorical missing data: A Comparison of multivariate normal and multinomial methods*. 20.11.2014 tarihinde <http://www.mwsug.org/proceedings/2008/stats/MWSUG-2008-S05.pdf> sitesinden alınmıştır.
- Gelman, A. ve Hill, J. (2012). Missing data imputation. *Data analysis using regression and multilevel/hierarchical models* (ss.529-544) içinde. Cambridge University Press.
- Ginkel, J.R.V., Van der Ark, L.A., Sijtsma, K. ve Vermunt, J.K. (2007). Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. *Computational Statistics & Data Analysis*, 51, 4013 – 4027, DOI:10.1016/j.csda.2006.12.022.
- Ginkel, J.R.V., Sijtsma, K., Van der Ark, L.A. ve Vermunt, J.K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. *Methodology*, 6 (1), 17-30, DOI: 10.1027/1614-2241/a000003
- Graham, J.W., Hofer, S.M. ve Piccinin, A.M.(1994). Analysis with missing data in drug prevention research. Collins, L.M. ve Seitz, L.A. (eds.). *Advances in data analysis for prevention intervention research* (ss. 13- 64) içinde. National Institutes of Health.
- Graham, J.W. (2009). Missing data analysis: Making it work in the real World. *Annual Review of Psychology*, 60, 549-576.
- Granberg- Rademacker, J.S. (2007). A comparison of three approaches to handling incomplete state level data. *State Politics and Policy Quarterly*, 7(3), 325-338.
- Haitovsky, Y. (1968). Missing data in regression analysis. *Journal of the Royal Statistical Society*, 30(1), 67-82.

- Hawthorne, G. ve Elliot, P (2005). Imputing cross-sectional missing data: comparison of common techniques. *Australian and New Zealand Journal of Psychiatry*, 39, 583-591, DOI: 10.1080/j.1440-1614.2005.01630.x.
- Hertel, B.D. (1976). Minimizing error variance introduced by missing data routines in survey analysis. *Sociological Methods & Research*,4, 459-474, DOI: 10.1177/004912417600400404
- Hohensinn, C. ve Kubinger, K.D. (2011). On the impact of missing values on the item fit and the model validness of the Rasch model. *Psychological Test and Assessment Modeling*, 53 (3), 380-393.
- Huck, S.W. (2012). *Reading statistics and research (6.Edition)*. USA: Pearson.
- Howell, D.C. (2007). The treatment of missing data. Outhwaite, W., Turner, S.P. (eds.), *The SAGE handbook of social science methodology* (ss. 208-225) içinde, Sage Publications.
- IBM (2012). 25.11.2014 tarihinde [http://www.01.ibm.com/support/knowledgecenter/SSLVMB\\_21.0.0/com.ibm.spss.statistics.help/syn\\_rmv\\_overview.htm](http://www.01.ibm.com/support/knowledgecenter/SSLVMB_21.0.0/com.ibm.spss.statistics.help/syn_rmv_overview.htm) sitesinden alınmıştır.
- IBM (2012). 28.11.2014 tarihinde [http://www-01.ibm.com/support/knowledgecenter/SSLVMB\\_20.0.0/com.ibm.spss.statistics.cs/mva\\_describe\\_rerun\\_mcartest.htm](http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.cs/mva_describe_rerun_mcartest.htm) sitesinden alınmıştır.
- Imtiaz, S.A. (2007). *The treatment of missing data in process monitoring and identification*. University of Alberta, Department of Chemical and Materials Engineering, Canada.
- Jing, L. (2012). *Missing data imputation*. Retrieved from UCLA Electronic Theses and Dissertations.

- Kalaycı, Ş. (Ed.) (2010). *SPSS uygulamalı çok değişkenli istatistik teknikleri* (5.Baskı). Ankara: Asil Yayın Dağıtım.
- Kang, H. (2013). The prevention and handling of missing data. *Korean J Anesthesiol*, 64(5), 402-406.
- Karal, Y. (2014). *Cox regresyon yöntemi modelinde kayıp veri analiz yöntemleri*. Yayınlanmamış Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Samsun.
- Karasar, N. (2007). *Bilimsel araştırma yöntemi: kavramlar, ilkeler, teknikler*. (17. Baskı). Ankara: Nobel Yayın Dağıtım.
- Kaspar, E.Ç. (2011). *Kayıp veriler ve kayıp veriler için bir çoklu veri atama yöntemi: Propensity skor*. Yayınlanmamış Doktora Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- Kenny, D.A. (1987). *Statistics for the social and behavioral science*. USA.
- Kim, J. ve Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Research*, 6(2), DOI: 10.1177/004912417700600206.
- King, G., Honaker, J., Joseph, A. ve Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1), 49-69.
- Köse, İ. A. ve Öztemur, B. (2014). Kayıp veri ele alma yöntemlerinin t-testi ve ANOVA parametreleri üzerine etkisinin incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 400-412.
- Lakshminarayan, K., Harp, S.A. ve Samad, T. (1999). Imputation of missing data in industrial databases. *Applied Intelligence* 11, 259–275.

- Larsen, R. (2011) Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(4), 649-662, DOI: 10.1080/10705511.2011.607721
- 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. ve Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Little, R.J.A. (1988). Missing data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6 (3), 287-296.
- McDonald, R.A., Thurston, P.W. ve Nelson, M.R. (2000). A monte carlo study of missing item methods. *Organizational Research Methods*, 3(1), 71-92, DOI: 10.1177/109442810031003.
- McKnight, P.E., McKnight, K.M., Sidani, S. ve Figueredo, A.J (2007). *Missing data: A gentle introduction*. United States of America: The Guilford Press.
- MEB (2011). *Pisa 2009 projesi ulusal ön raporu*. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Musil, C.M., Warner, C.B, Yobas, P.K. ve Jones, S.L. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, 24(7), 815-829, DOI: 10.1177/019394502237390.
- Newman, D.A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, 6(3), 328-362, DOI: 10.1177/1094428103254673

- Oğuzlar, A. (2001, Eylül). *Alan arařtırmalarında kayıp deęer problemi ve çözümler önerileri*. V. Ulusal Ekonometri ve İstatistik Sempozyumu'nda sunulan bildiri. Çukurova Üniversitesi, Adana.
- Olinsky, A., Chen, S. ve Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, 151, 53–79, DOI:10.1016/S0377-2217(02)00578-7.
- Osborne, J.W. (2013). *Best practices in data cleaning*. New York: Sage Publications.
- Peng, C.-Y. J., Harwell, M., Liou, S.-M., Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (ed.), *Real data analysis* (ss.31-78) içinde. Greenwich, CT: Information Age Publishing Inc.
- Peugh, J.L. ve Enders, J.K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525-556, DOI: 10.3102/00346543074004525.
- Pigott, T.D. (2001). A review of methods for missing data. *Educational Research and Evaluation*, 7(4), 353-383.
- Pokropek, A. (2011). Missing by Design: Planned missing-data designs in social science. *Research & Methods*, 20 (1), 81–105.
- Rose, R.A. ve Fraser, M.W.(2008). A simplified framework for using multiple imputation in social work research. *Social Work Research*, 32(3), 171-178.
- Roth, P.L. (1994). Missing data: A conceptual review for applied psychologist. *Personnel Psychology*, 47(3), 537-560.
- Roth, P.L. ve Switzer, F.S. (1995). A monte carlo analysis of missing data techniques in a HRM setting. *Journal of Management*, 21(5), 1003-1023.

- Roth, P.L., Campion, J.E. ve Jones, S.D. (1996). The Impact of Four Missing Data Techniques on Validity Estimates in Human Resource Management. *Journal of Business and Psychology*, 11(1), 101-112.
- Roth, P.L., Switzer III, F.S. ve Switzer, D.M. (1999). Missing data in multiple item scales: A monte carlo analysis of missing data techniques. *Organizational Research Methods*, 2(3),211-232, DOI: 10.1177/109442819923001.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse surveys*. USA: John Wiley & Sons, Inc.
- Satıcı, E. ve Kadılar, C. (2009). Kayıp gözlem olduğunda kitle ortalamasının tahmini. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 10 (2), 549-556.
- Saunders, J.A., Howell, N.M., Spitznagel, E., Dore, P., Proctor, E.K. ve Pescarino, R. (2006). Imputing missing data: A comparison of methods for social work researchers. *Social Work Research*, 30(1), 19-31.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. USA: Chapman & Hall.
- Schafer, J.L. ve Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177, DOI: 10.1037//1082-989X.7.2.147.
- Schlomer, G.L. Bauman, S. ve Card, N.A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, 57(1), 1–10, DOI: 10.1037/a0018082.
- Sezgin E. ve Çelik Y.(2013). *Veri madenciliğinde kayıp veriler için kullanılan yöntemlerin karşılaştırılması*, Akademik Bilişim Konferansı (23-25 Ocak 2013), Akdeniz Üniversitesi.

- Sinharay, S., Stern, H.S. ve Russel, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, 6 (4), 317-329.
- 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.
- Suraphee, S., Raksmanee, C., Busaba, J., Chaisorn, C. ve Nakornthai, W. (2006). A comparison of estimation methods for missing data in multiple linear regression with two independent variables. *Thailand Statistician*, 4,13-26.
- Şirin, S.R. ve Vatanartıran, S. (2014). *PISA 2012 değerlendirmesi: Türkiye için veriye dayalı eğitim reformu önerileri*. İstanbul: TÜSİAD.
- Tabachnick, B. G. ve Fidell, L.S. (2007). *Using multivariate statistics (5th Edition)*. USA: Pearson.
- Tabachnick, B. G. ve Fidell, L.S. (2013). *Using multivariate statistics (6.Edition)*. USA: Pearson.
- Tan, Ş. (2012). *Öğretimde ölçme ve değerlendirme: KPSS el kitabı (7.Baskı)*. Ankara: Pegem Akademi.
- Tekin, H. (2012). *Eğitimde Ölçme ve Değerlendirme (21. Baskı)*. Ankara: Pegem Akademi.
- Van der Ark, L. A., ve Vermunt, J. K. (2010). New developments in missing data analysis. *Methodology*, 6 (1), 1-2, DOI: 10.1027/1614-2241/a000001
- Vansteelandt, S., Carpenter, J. ve Kenward, M.G. (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, 6 (1), 37-48. DOI: 10.1027/1614-2241/a000005.

Yılmaz, H. (2014). *Random forests yönteminde kayıp veri probleminin incelenmesi ve sağlık alanında bir uygulama*. Yayımlanmamış Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.

Young, W., Weckman, G. ve 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, DOI: 10.1080/14639220903470205

Wayman, J.C. (2003). *Multiple imputation for missing data: What is it and how can I use it?* Annual Meeting of the American Educational Research Association'da sunulmuş bildiri. Chicago, IL.

Wilkinson, L. ve The Task Force on Statistical Inference APA Board of Scientific Affairs (1999). *American Psychologist*, 54(8), 594-604.

## ÖZGEÇMİŞ

1988 Ankara doğumludur. İlk ve orta öğrenimini Ankara'da tamamladı. 2006-2011 yılları arasında Orta Doğu Teknik Üniversitesi Fen Bilgisi Öğretmenliği bölümünde üniversite eğitimini tamamladı. 2012 yılında O.D.T.Ü.'de Fen Bilgisi Öğretmenliğinde yüksek lisansa başladı. Üniversite'den mezun olduktan sonra 1 yıl özel bir eğitim kurumunda çalıştı. 2013 yılında Abant İzzet Baysal Üniversitesi, Eğitimde Ölçme ve Değerlendirme bölümünde araştırma görevlisi olarak göreve başladı.