

T.C.

HASAN KALYONCU ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANA BİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME TEZLİ YÜKSEK LİSANS PROGRAMI

**ABİDE-2016 MATEMATİK TESTİNİN FARKLI KAYIP VERİ TEKNİKLERİ İLE
İNCELENMESİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN

ABDURRAHMAN KAYA

GAZİANTEP - 2022

T.C.

HASAN KALYONCU ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANA BİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME TEZLİ YÜKSEK LİSANS PROGRAMI

**ABİDE-2016 MATEMATİK TESTİNİN FARKLI KAYIP VERİ TEKNİKLERİ İLE
İNCELENMESİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN

ABDURRAHMAN KAYA

TEZ DANIŞMANI

DOÇ.DR. UFUK AKBAŞ

GAZİANTEP - 2022

TEZ ETİK VE BİLDİRİM SAYFASI

Yüksek lisans tezi olarak sunduğum “**ABİDE-2016 Matematik Testinin Farklı Kayıp Veri Teknikleri ile İncelenmesi**” başlıklı çalışmanın tarafımda, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu ve bunlara atıf yapılarak yararlanmış olduğumu belirtir ve onurumla doğrularım.

Abdurrahman KAYA



ÖNSÖZ

Bu arařtırmada ABİDE-2016 Matematik Testinin A kitapçığı Çoktan seçmeli maddelerinin kayıp verilerinin Klasik Test Kuramına göre farklı kayıp veri yöntemlerine göre psikometrik özellikleri istatistiksel analiz teknikleri ile incelenmiştir. Bulunan sonuçlar ile ABİDE-2016 yılında uygulanan matematik testinden elde edilen bulgular ve bulgulara dayalı olarak önerilerde bulunulmuştur.

Yaptığım çalışmada bana her konuda destek olan, bana yol gösteren, yüksek lisans ve tez dönemlerinde hoşgörölü ve nazik yaklaşımıyla çalışmamın her aşamasını dikkatle inceleyip gerekli düzenlemeleri yapmamı sağlayan tez danışmanım ve çok değerli hocam Dr.Öğr. Üyesi Ufuk AKBAŞ'a, öğrenimim süresinde bana kattıklarından dolayı başta Prof. Dr. Şener BÜYÜKÖZTÜRK olmak üzere tüm hocalarıma ve tez sürecimde her aşamada bana yardımını hiç esirgemeyen Araştırma görevlisi Merve YILDIRIM SEHERYELİ'ne sonsuz teşekkürlerimi sunarım. Ayrıca verilere ulaşmamda yardımcı olan Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü'ne teşekkür ederim.

Akademik hayatım boyunca hep yanımda olan beni destekleyen, sevgisini hiçbir zaman eksik etmeyen annem Fatma KAYA, babam Mehmet KAYA'ya ve kardeşlerime teşekkür ederim. Yüksek lisans eğitimimde beni sürekli motive eden manevi desteğini yanımda hissettiğim eşim Zeynep KAYA'ya, ve neşe kaynağım olan oğlum Ali Eymen KAYA ve Mehmet Talha KAYA'ya teşekkür ederim.

Gaziantep, 2022

Abdurrahman KAYA

ÖZET

Bu çalışmada ABİDE-2016 sınavının 8. sınıf öğrencilerine uygulanan Matematik Testi A kitapçığının, iki kategorili çoktan seçmeli dokuz maddesinin psikometrik özelliklerinin kestiriminde, farklı kayıp veri yöntemlerine göre uygun olan kayıp veri yönteminin belirlenmesi amaçlanmaktadır.

Veri analizleri yanıtlayıcılar tarafından cevaplanan 1-0 şeklindeki iki kategorili veri seti üzerinden 4 farklı kayıp veri yöntemi kullanılarak yapılmıştır. Yapılan bu analizler Klasik Test Kuramına dayalı çalışmalardır. Kayıp veri yöntemi olarak Liste Bazında Silme, Hot-Deck Atama, Stokastik Regresyonla Atma ve Çoklu Değer Atama yöntemleri kullanılmıştır.

Yapılan analizler sonucunda eldeki bulgular, iki kategorili puanlanan maddelerden oluşan veri setleri için seçkisiz olmayan kayıp veriler için uygun kayıp veri tekniklerinin kullanılmasının gerekliliğini göstermektedir.

Anahtar Kelimeler: Kayıp veri, Abide sınavı, Klasik test kuramı

ABSTRACT

In this study, it is aimed to determine the missing value method according to different missing value methods in estimating the psychometric properties of the dichotomous multiple-choice 9 items of the Mathematics Test A booklet applied to the 8th grade students of the ABIDE-2016 exam.

Value analyzes were carried out using 4 different missing value methods on a dichotomous data set, 1-0, answered by the respondents. These analyzes are based on Classical Test Theory. Listwise Deletion, Hot-Deck Imputation, Stochastic Regression Imputation and Multiple Imputation methods were used as missing value methods.

As a result of the analysis, the findings show the necessity of using appropriate missing value techniques for not missing at random value for data sets consisting of items scored in two categories.

Keywords: Missing value, Abide Exam, Classical Test Theory

İÇİNDEKİLER	Sayfa No.
ÖNSÖZ	i
ÖZET	ii
ABSTRACT	iii
İÇİNDEKİLER	iv
TABLolar LİSTESİ	vi
ŞEKİLLER LİSTESİ	Vii
KISALTMALAR	viii

BİRİNCİ BÖLÜM

GİRİŞ	1
1.1. Problem Durumu.....	1
1.1.1. Problem Cümlesi.....	9
1.1.2. Alt Problemler.....	9
1.2. Araştırmanın Amacı.....	10
1.3. Araştırmanın Önemi.....	10
1.4. Sayıtlar.....	10
1.5. Araştırmanın Sınırlılıkları.....	11

İKİNCİ BÖLÜM

KAVRAMSAL ÇERÇEVE	11
2.1. Ölçme ve Değerlendirme.....	11
2.1.1. Ölçme	11
2.1.2. Eğitimde Ölçme	12
2.2. Akademik Becerilerin İzlenmesi ve değerlendirilmesi (ABİDE).....	13
2.2.1. ABİDE Araştırmasının Genel Amacı	15
2.2.2. ABİDE-2016 Araştırmasının Yöntemi	16
2.3. Kayıp veri Mekanizmaları.....	18
2.3.1. Kayıp Veri Sürecinde Rastgeleliğin Sorgulanması.....	26
2.3.2. Kayıp Veri İle Baş Etme Yöntemleri.....	26
2.3.2.1. Silme Yöntemleri.....	27
2.3.2.2. Atama Yöntemleri.....	29
2.4. Klasik Test Kuramı.....	34
2.4.1. Klasik Test Kuramında Madde istatistikleri.....	36
2.4.2. Klasik Test Kuramı'nın Avantajları ve Dezavantajları.....	42
2.5. İlgili Araştırmalar.....	44

ÜÇÜNCÜ BÖLÜM

YÖNTEM	47
3.1. Araştırma Modeli.....	48
3.2. Evren ve Örneklem.....	48
3.3. Veri Toplama Araçları.....	48
3.4. İşlem	49
3.4. Verilerin Analizi ve Yorumlanması.....	50

DÖRDÜNCÜ BÖLÜM

BULGULAR VE YORUM	51
4.1. Kayıp Veri Örüntüsü ve Mekanizması.....	51
4.2. Test ve Madde Parametreleri.....	57

BEŞİNCİ BÖLÜM

SONUÇLAR VE ÖNERİLER

5.1. Sonuçlar.....	62
5.2. Öneriler.....	64
KAYNAKÇA	65

TABLolar LİSTESİ

Sayfa No.

Tablo 1. Üç Farklı Kayıp Veri Mekanizması İçin Örnek Veri Seti.....	21
Tablo 2. Kayıp veri ile ilgili değişkenlerin analize dahil edilme durumu.....	25
Tablo 3. Liste Bazında ve Çift Bazında Silme Tekniklerinin Karşılaştırılması.....	28
Tablo 4. Test Formları ve Madde Sayıları.....	48
Tablo 5. Kayıp verilerin Cinsiyet'e Göre Frekans Dağılımı.....	52
Tablo 6. Kayıp Verilerin Cinsiyete Göre Dağılımı.....	55
Tablo 7. . Kayıp Sayısına Göre Madde Ortalamalarının Karşılaştırılması.....	55
Tablo 8. Kayıp Veri Yöntemlerine Göre Madde ve Test parametreleri.....	58

ŞEKİLLER LİSTESİ

Sayfa No.

Şekil 1. Tamamen Seçkisiz Kayıp Mekanizmasına ilişkin Gösterim	22
Şekil 2. Seçkisiz Kayıp Mekanizmasına ilişkin Gösterim.....	23
Şekil 3. Seçkisiz Olmayan Kayıp Mekanizmasına ilişkin Gösterim.....	24
Şekil 4. Çoklu Değer Atama İşlem Adımları.....	34
Şekil 5. Araştırmada İşlenen İşlem Yolu.....	49
Şekil 6. Kayıp Veri Örüntü Grafiği.....	53
Şekil 7. Kayıp Veri Örüntülerine İlişkin Frekans Dağılımı.....	54
Şekil 8. Toplam Puan ile Kayıp Sayısı Arasındaki Saçılma Grafiği.....	56

KISALTMALAR LİSTESİ

ABİDE:	Akademik Becerilerin İzlenmesi ve Değerlendirilmesi
MEB:	Millî Eğitim Bakanlığı
OECD:	Organization for Economic Cooperation and Development
ÖSYM:	Ölçme, Seçme ve Yerleştirme Merkezi
PISA:	Program for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)
TIMSS:	Trends International Mathematics and Science Study (Uluslararası Matematik ve Fen Eğilimleri Araştırması))
PIRLS :	Progress in International Reading Literacy Study (Uluslararası Okuma Becerileri Projesi)
OECD :	Ekonomik İşbirliği ve Kalkınma Teşkilatı
MEB :	Milli Eğitim Bakanlığı
AFA :	Açımlayıcı Faktör Analizi
LW :	Listwise (Liste Bazında Silme)
HD :	Hot-Deck Atama
SRA :	Stokastik Regresyon ile Atama
ÇDA :	Çoklu Değer Atama
KTK :	Klasik Test Kuramı
MTK :	Madde Tepki Kuramı
TSK :	Tamamen Seçkisiz Kayıp
SK :	Seçkisiz Kayıp
SOK :	Seçkisiz Olmayan Kayıp
KMO :	Kaiser-Meyer-Olkin
p :	Madde Güçlük İndeksi
r _{jx} :	Madde Ayırt Edicilik Gücü
N :	Kişi Sayısı
X :	Aritmetik Ortalama
ss :	Standart Sapma
sd :	Serbestlik Derecesi
p, α :	Anlamlılık Düzeyi

BİRİNCİ BÖLÜM

GİRİŞ

1.1. Problem Durumu

Değişen dünyamızda eğitim, bilim ve teknoloji gibi alanlarda ülkeler hızlı değişim ve gelişme göstermektedir. Ülkeler bu gelişmelere ayak uydurabilmek ve çağın beklentilerine cevap bulabilmek için eğitim sisteminin geliştirmeye çalışmaktadır. Ülkelerin kendilerini geliştirmedeki en büyük adım ise eğitimidir. Bu nedenle ülkelerdeki eğitime verilen önem gün geçtikçe artmaktadır. Değişen dünyamızda dünyayı daha yakından tanımak ve ayak uydurabilmek için, araştırma ve sorgulamayı seven, aldığı eğitimi günlük hayat ile bağdaştırabilen, öğrendiklerini günlük hayata transfer edebilen özgüven duygusu yüksek bireyler yetiştirmek için eğitim sisteminde değişikliklere gitmektedir.

Günümüz dünyasında gelişmiş ve gelişmekte olan ülkelerin ekonomilerinin gelişmesine katkı sağlayan, toplum refah düzeyinin gelişmesine öncülük eden, kendi kültürlerinin özümsemesini sağlayan ve konuda çaba sarf eden, diğer toplumlarla ve kendi toplumuyla iletişimde empati kurabilen insan gücü yetiştirmek ve kendini geliştiren bireylerin yetişmesi ancak eğitim kalitesinin artırılması ile gerçekleştirilebilir. Bu sebeple ülkelerin eğitimin geliştirilmesi için gerekli çalışmaları yapması gerekmektedir (MEB,2009).

Eğitimin-öğretim programları sonucunda elde edilen verilerin yorumlanması ve bu yorumlamalar neticesinde ulaşılan veriler ile öğretim programlarının geliştirilmesine yardımcı olunması gerekir. Bu sebeple eğitim sisteminin daha etkili olması için yapılan değerlendirmelere büyük önem vermek gerekmektedir. Okul hayatının başlamasından itibaren öğrencilerin eğitim öğretim hayatındaki davranış ve öğretim durumlarının takip edilip eğitim sisteminin hedefine ulaşıp ulaşılmadığını görmek büyük önem arz etmektedir.

Okul hayatının hedefleri öğrenim programlarının gerçekleşmesi önemli bir yer tutar tutmaktadır. Dolayısıyla iyi bir ölçme ve değerlendirme sisteminin öğretim programının hedeflerinin hangi oranda gerçekleştirdiğini ve bu hedeflerin istenilen seviyeye ulaşmasına katkı sağlaması gerekmektedir. Kutlu, Doğan ve Karakaya (2014), okul hayatının eğitim ve öğretim hedefleri doğrultusunda da günlük hayatla bağdaştırılması büyük önem arz etmektedir.

Ölçme ve değerlendirme, öğrenme sürecinin ve öğretim programlarının önemli bir ögesidir. Sadece okul hayatındaki başarıyı ölçmesi değil aynı zamanda bir bütün olarak eğitim-öğretim hayatını da kapsayan bir ölçme ve değerlendirme programı önem arz etmektedir. Etkililiği kanıtlanmış olan bir ölçme ve değerlendirme programı, eğitim-öğretim hayatında kanıta dayalı daha net sonuçların alınacağı ve eğitim-öğretimin önemli bir bileşeni durumuna getirecektir (MEB,2017).

Etkili sonuçlar veren ölçme programlarının, öğretim yöntemlerinde ve değerlendirme aşamasındaki ölçme araçlarındaki değişimi de beraberinde getirmektedir. Öğrencilere kazandırılmak istenen davranışların ne kadarının öğretilip ve hangi seviyede olup olmadığını anlamak için bir ölçme ve değerlendirmenin olması gerekmektedir. Bu nedenle ülkeler eğitim sistemlerinin geliştirilmesi için geri bildirim vermek ve yapılan çalışmaların istenilen hedeflere ulaşp ulaşmadığını belirlemek için uluslararası sınavlar yapmaktadır. Türkiye, ulusal düzeyde yapılan sınavların yanında aynı zamanda uluslararası yapılan sınavlara katılım göstermektedir.

Milli Eğitim Bakanlığı kurucusu olduğu Ekonomik İşbirliği ve Kalkınma Örgütü'nün (Organisation for Economic Co-operation and Development:OECD) eğitimin geliştirilmesi ve öğretim uygulamalarının değerlendirilmesi için uluslararası çalışmalar yapmaktadır. Bu çalışmalar, ülkeler arası karşılaştırmanın yapılmasından ziyade sınavların yapılmasının genel amacının ülkelerin eğitim sistemlerindeki eksiklikleri görmeleri ve bu çerçevede değişiklik yapmalarına olanak sağlamaktadır. Bu çalışmaları yapan kuruluşlar, Uluslararası Eğitimsel Başarıyı Değerlendirme Birliği (International Association for the Evaluation of Educational Achievement:IEA), Dünya Bankası (World of Bank:WB) ve Birleşmiş Milletler Eğitim, Bilim ve Kültür Örgütü (United Nations Educational, Scientific and Cultural Organization) UNESCO bulunmaktadır. Bu kuruluşların yaptığı eğitim çalışmalarından bazıları, Uluslararası Okuma Becerilerini Geliştirme Çalışması (The Progress in International Reading Literacy Study: PIRLS), Uluslararası Matematik ve Fen Eğilimleri Çalışması (Trends in International Mathematics and Science Study: TIMSS) ve Uluslararası Öğrenci Değerlendirme Projesi (Program for International Student Assessment: PISA) dır (MEB, 2005).

Milli eğitim bakanlığı tarafından ulusal düzeyde yapılan öğrenci başarısını ölçme ve değerlendirme çalışmalarının dışında uluslararası alanda da, öğrencilerin

seviyelerini ve eğitim sistemimizi farklı ülkelerin sistemleri ile karşılaştırarak güçlü ve eksik yönlerini belirlemek için 2003 yılında, kurucu üyesi olduğu Ekonomik İşbirliği ve Gelişme Örgütü'nün (Organisation for Economic Co-operation and Development:OECD), kısa adı PISA olan Uluslararası Öğrenci Değerlendirme Projesi'ne (Program for International Student Assessment) katılmıştır (MEB, 2005). Türkiye, 1999'dan itibaren (2003 yılı hariç) TIMSS'e, 2003 yılından bu yana da PISA'ya katılıyor. Her iki sınavdan alınan veri sonuçları ile ülkemizin kendi içerisinde değerlendirilmesi ve diğer ülkeler ile karşılaştırılması bakımından büyük önem arz etmektedir. Ülkemizde öğrencilerin istedik davranışların elde edilip edilmediğini belirlemek için ulusal ve uluslararası düzeydeki sınavlar ile değerlendirilmeye çalışılmaktadır. Uluslararası düzeyde yapılan sınavlardan PISA, TIMSS ve PIRLS sınavlarında değerlendirme sonucunda öğrenme çıktılarının düşük seviyede olduğu belirlenmiştir (MEB, 2005). Bu görüşü destekler nitelikte, Gelbal ve Kelecioğlu (2007), Ülkemizdeki eğitim hayatı boyunca öğrendikleri bilgilerin günlük hayatta karşılaştıkları problem ve sorunların çözümünde kullanmalarında eksiklikleri olduğu görülmektedir. Demirtaşlı (2014), bu eksikliklerin giderilmesi için okul eğitiminde verilen derslerdeki öğrenmelerin geri dönütü için belirli aralıklarla standart ölçme ve değerlendirme araçlarına gereksinim vardır. Ülke çapında yapılan sınavların geri bildirim almak için ölçme değerlendirme çalışmalarının yeterli olmadığı görülmektedir. Ülkemizde sadece öğrencileri sıralamaktan ibaret olan sınavlardan ziyade geçerliliği ve güvenilirliği yüksek ölçme ve değerlendirme uygulamalarına gereksinim vardır. Ayrıca eğitim sistemimizin eksikliklerinin farkına varılmasına yönelik analizlerin yapılabilmesi için, öğrencilerin başarı ve beceri düzeylerinin farklı ülkelerdeki öğrenci başarıları ile mukayese edilmesi ülke içerisindeki değerlendirmeler kadar önem arz etmektedir.

Milli Eğitim Bakanlığı 2015-2019 ve 2019-2023 Stratejik planında eğitim-öğretim sürecinde öğrencilerin, zamanın ihtiyaç duyduğu bilgi ve becerilerin kazandırılması için eğitim-öğretim sistemimizin geliştirilmesi gerektiği vurgulanmaktadır (MEB,2015 ve MEB,2019). Dünya genelinde yenilikçi uygulamalara ve günümüz standartlarına önem veren ülkeler ölçme ve değerlendirme programlarına önem vermektedir (Stiggins, 1999).

Eđitim sistemimizde 2005-2006 yılından itibaren benimsenmiř olan đretimden ziyade đrenme odaklı bir sistem olan “yapılandırmacı yaklařım” modeli sistemi gz nnde bulundurularak ders programlarında, ders ieriklerinde deđiřiklikler yapılmaya alıřılmıřtır. Bu yaklařımla birlikte dnya zerindeki deđiřim ve geliřmeleri eđitim sistemine yansıtılmak iin yeni eđitim mfredatları geliřtirilmeye alıřılmıřtır. Milli Eđitim Bakanlıđı bnyesindeki Talim Terbiye Kurulu Bařkanlıđı yapılandırmacı yaklařımın benimsenmesindeki asıl sebebin deđiřen ve geliřen dnyamızda bilgiyi retebilen rettiđi bilgiyi kullanabilen, arařtırmayı seven, eleřtirel dřnebiyen, karřılařtıđı problemler karřısında zm yolları reten, rettiđi zm yollarını uygulayabilen ve karar alma becerileri geliřmiř bireyler yetiřtirmektedir. (MEB, 2009). lkemizdeki eđitim sistemindeki grlen eksiklikler, ulusal ve uluslararası yapılan sınavlardan sonra elde edilen sonulara bakılarak eđitim sistemimizde bazı eksiklikler veya aksaklıklar olduđu bu sebeple MEB bakanlıđı bazı yıllarda sınav deđiřikliđine gitmiřtir.

MEB, 2005-2006 yılından itibaren 8. Sınıf sınav sistemlerinde bazı deđiřiklik yapmaya bařlamıřtır. LGS (Liselere Geiř Sınavı), 2005 yılında OKS (Orta đretim Kurumları Sınavı) geiř, 2008 yılında SBS (Seviye Belirleme Sınavı) geiř, ve son olarak 2013 yılında TEOG (Temel đretimden Orta đretime Geiř Sınavı) kaldırılmıř yerine 2017 yılından itibaren LGS(Liseye Giriř Sınavı) getirilmiřtir. Deđiřim gsteren eđitim sistemimizle birlikte bir eđitim seferberliđi olduđu gzlenmektedir. Eđitim-đretim sistemindeki bu kadar yapılan deđiřim ve yeniliklerle okuldaki ders iřlenme sisteminin deđiřimini, ulusal ve uluslararası sınavları etkileyeceđi bir gerektir (Eđitim Reformu Giriřimi,2013; Atılgan,2017; Kaplan,2017; Ekinci, O. ve Bal, Ayten P.2019 ve diđerleri).

Yapılan bu sınavların nemi kadar bu sınavların lme ve deđerlendirme kriterlerine uyumu da eđitim ve đretim sonucunda dođru bilgilere ulařmak iin nemlidir (Wu, Li ve Zumbo, 2007). Tm bu yapılan arařtırmalar ve alıřmalar gz nne alındıđında Milli Eđitim Bakanlıđı lme ve Deđerlendirme ve Sınav Hizmetleri Genel Mdrlđ (DSGM) đretim programlarının hedeflerini gzeterek, okul hayatında ne đrendiklerinden ziyade gnlk hayata ne kadar aktardıkları ve neler yapabildikleri zerinde durmaktadır. Bu nedenle đrencilerin st dzey đrenme becerilerinin lcldđ ulusal dzeyde bir izleme ve deđerlendirme alıřması iin geliřtirilen Akademik Becerilerin İzlenmesi ve Deđerlendirilmesi (ABİDE) adı altında 8. sınıf đrencilerine matematik, Trke, fen bilimleri ve sosyal

bilimler alanlarında uygulanan bir sınav yapmıştır. Ulusal sınavlar eğitim sisteminin geliştirilip ve değerlendirilmesi için ilk öncelik olarak ele alınmaktadır (MEB, 2017c). Türkiye'nin de PISA ve TIMSS gibi üst düzey düşünme becerilerini ölçmek için ABİDE programını uygulamaya koyması eğitim-öğretim ile ilgili araştırmalara önem verdiğini göstermektedir.

Dünyadaki birçok ülkede matematik öğretimin öğrenilmesinden ziyade, bu öğrenimin nasıl olması konusunda birbirinden farklı yaklaşımlar göstermektedir. Son yıllarda birçok ülkede matematik öğretiminde matematiği gerçek yaşam şartlarında uygulamak ve bilimsel çalışmalarda kullanma gibi stratejiler izlemektedir (MEB,2016b). Matematik öğretiminde çağa ayak uydurulacak yeni düzenlemeler yapmak bir zorunluluk oluşturmuştur. Dünyadaki tüm bu gelişmeler ışığında matematik okuryazarlığı büyük önem kazanmaktadır. Bu sebeple 2006-2007 yılından itibaren ülkemizde matematik derslerinde yapılandırmacı bir yaklaşımla öğretim programları uygulandığı ve hızla gelişen dünyamızda matematiğe verilen önemin gün geçtikçe arttığı gözlenmektedir (Ersoy, 1997).

PISA ve TIMSS sınavlarında öğrencilerin öğrendiklerinden nasıl anlam çıkardıkları ve matematik becerilerini gerçek yaşamdaki durumlarda nasıl kullanacaklarını değerlendirmeyi amaçlar (MEB,2015). PISA ve TIMSS gibi sınavlarda matematiğe verilen önem gibi ülkemizde de ABİDE sınavında matematiğe önem verildiği yapılan çalışmalardan anlaşılmakta ve ölçülmektedir. Ulusal ve uluslararası düzeyde yapılan tüm bu sınavlarda matematik ve diğer derslerin öğretiminin etkili sonuç verip vermediğini anlamak için ölçme ve değerlendirme çalışmalarında farklı bilişsel alana yönelik başarı testlerinin kullanılması da önem kazanmaktadır. Dolayısıyla başarı testlerinde yer alan madde formatları, ölçülmek istenen özelliğe göre değişmektedir. ABİDE, PISA ve TIMSS gibi sınavlarda da bu formatların kullanıldığı gözlenmektedir. Türkiye'de bilişsel alanda uygulanan ölçme ve değerlendirmelerin çoğunluğunu başarı testleri oluşturmaktadır.

Testlerde kullanılan soru tipleri ve madde formatları, başarı testlerini diğer veri toplama araçlarından ve kendi içerisinde birbirlerinden ayıran en önemli özelliklerdir. Öğrencilerin farklı yetenek düzeyleri ve farklı öğrenme biçimlerine sahip olmaları sebebiyle ve değerlendirmelerinde farklı amaçlarla yapılmasından dolayı farklı ölçme ve değerlendirme test formatlarının kullanılmasını gerektirmektedir (Baykul,2000).

Testlerde yer alan soru türlerinin ve madde formatlarının istediğimiz amaca uygun, istenilen davranışı ölçebilecek yeterlilikte ve özellikte olmalı ve istenilen bir davranışı

ölçmek için en uygun madde formatının belirlenmesi önem arz etmektedir. Aksi durumda kullanılan madde formatı ölçmek istediğimiz özelliği ve davranışı ölçemeyecek veya farklı sonuçlar çıkmasına sebep olacak ve geçerliliğin düşük olmasına yol açacaktır. Bu nedenledir ki geniş kapsamlı sınavlarda, aynı soru türü üzerinde ısrar etmek yerine farklı soru tarzlarının kullanılmasına da olanak sağlanmalıdır (Huba ve Freed, 2000).

Eğitim-öğretim programlarında ölçme ve değerlendirme çalışmalarında genellikle çoktan seçmeli testlerle, açık uçlu sorular ve doğru-yanlış testleri uygulanmaktadır. Fakat bu ölçme ve değerlendirme çalışmalarıyla ilgili farklı tartışmalar ve araştırmalar bulunmaktadır (Baki ve Birgin, 2002; Francisco, Nakhleh, Nurrenbern ve Miller, 2002). Bu tartışmalar ile ilgili literatür taraması yapıldığında her test formatının eksi ve artı yönlerinin olduğu anlaşılmaktadır. Yapılan ulusal ve uluslararası sınavlar hazırlanırken hangi düzeyde yeteneği ölçmek isteniyor ise o düzeye uygun sınav formatları hazırlanmaktadır. Haladyna'ya (1997) göre eğitimde yapılan ölçmede kullanılan farklı madde türlerinin güçlü ve zayıf yönlerinin olduğunu belirtmektedir. Çoktan seçmeli maddelerin diğer madde türlerine göre puanlama kolaylığı, objektifliği ve kısa sürede daha fazla beceri ölçebilmesinden dolayı üstünlüğü bulunmaktadır. Fakat bu üstünlüğünün yanı sıra bazı dezavantajları da bulunmaktadır. Tekin'e (2017) göre çoktan seçmeli testlerin tesadüfen doğru yanıt bulma olasılığı yüksektir. Bu şans başarısı çoktan seçmeli testler için geçerlilik ve güvenilirliği olumsuz etkilemektedir.

Farklı yetenekli öğrencilerin değerlendirilmeye alınmasında dolayı PISA ve TIMSS gibi sınavlarda çeşitli bilişsel testler kullanılmaktadır. Bu bilişsel testler öğrenciyi çok boyutlu değerlendirmektedir. Çok boyutlu değerlendirmeyi yapabilmek için PISA ve TIMSS'te kullanılan madde formatları; çoktan seçmeli, karmaşık çoktan seçmeli, açık uçlu ve kısa cevaplı sorular uygulanmaktadır (MEB,2016b). ABİDE programında ise bu madde formatlarından çoktan seçmeli test ve açık uçlu madde formatı kullanılmaktadır (MEB,2017c). Farklı tarzda soru kullanılması açık ve uzaktan eğitimde sıklıkla kullanılan bir yoldur. Sınavların ayırt ediciliklerini ve güçlüklerini sınavdan önce bilmek, yapmak istediğimiz sınavı kontrol altında tutmamıza yardımcı olacaktır (Akyıldız ve Karadağ,2018).

Tüm bu eğitim ve psikolojideki değişkenlerin ölçülmesi aslında gözlenemeyen bir gizil değişkeni ölçmekle alakalıdır. Aslında bu gizil değişkenlerin ölçülmesi, gizil değişkenin kendisini yansıttığını düşünülen gözlenen değişkenler yardımıyla

ölçülmektedir. Kişilerin bu gözlenen değişkenlere cevap vermesi aslında ölçülmek istenen gizil değişkene verdiği cevap olarak yorumlanmaktadır.

Bu gizil değişkenlerin ölçülmesi ile ilgili çeşitli ölçme ve değerlendirme araçları geliştirilmiştir. Bu ölçme araçlarının geliştirilmesi ölçmek istenilen değişkenler ile ilgili hata yapılmaması için ölçme araçlarının daha dikkatli ve titiz hazırlanması gerekmektedir. Çünkü ölçme yapan kişinin, ölçme aracının geçerlilik ve güvenilirliğini sağlamak, sorgulamak araştırmacının önemli sorumluluklarından biridir (Akbaş, 2014).

Araştırmacı geçerlilik ve güvenilirlik ile ilgili kişisel yorumlarını katmadan yani nesnel sonuçlar elde etmek için farklı istatistik analizlere başvurmaktadır. Bu istatistiksel programlar ile yapılan veri analizleri arasında farklı teknik özellikler bulunmaktadır. Fakat bu istatistiksel programlar ölçülmek istenen değişkenin eksiksiz olduğunu yani herhangi bir veri kaybı olmadığı üzerine kuruludur. Bu kayıp veriler ölçülmek istenen değişkenler ile ilgili elde edilen bulguların yorumlanmasını sınırlandırmaktadır. Dolayısıyla elde edilen bu bulgular istenilen hedefe ulaşmada eksik kalacaktır.

Eksiksiz veriler için hazırlanan programlar nedeniyle yapılan analizler sonucunda elde edilen kestirimlerin istatistiksel gücün düşük veya yüksek çıkmasına, sonuçların geçerlilik ve güvenilirliğini, standart hatanın artması, ortalama ve standart sapma vb. gibi istatistiksel değerleri olumsuz yönde etkileyecektir (Little ve Rubin, 1987; Acock, 2005).

Kayıp veri, ölçülmek istenen özelliğe yönelik elde bir gözlemin olmaması durumu olarak tanımlanmaktadır (Hohensinn ve Kubinger, 2011). Kayıp veriler elde edilen veri setinin küçülmesi ve yapılacak kestirimlerin güçsüz olması anlamına gelmektedir (Rubin, 1987).

Literatür taraması yapıldığında kayıp veriler ve kayıp verilerin sebepleri ile ilgili farklı açıklamalar yapılmaktadır. Araştırmaya katılan bireylerin ölçme aracındaki bir veya birden fazla maddeye ait verilerinin olmaması madde yanıtlamama (item nonresponse); araştırmaya katılan bireye ulaşılamaması veya araştırmaya katılan kişilerin maddelere cevap vermemesi gibi durumlar birim yanıtlamama (unit nonresponse) olarak adlandırılmaktadır. Maddelerin yanıtlanmaması durumunda veri setindeki bazı hücrelerin boş kalmasına, birim yanıtlamama durumunda ise araştırmaya katılan bireyler hakkında bilgi olmaması göstermektedir (Heerwegh, 2005; De Luca ve Perrachi, 2007).

1970 yılına kadar kayıp verilerin istatistiksel analizlerde bir sorun olarak görülmediği yirminci yüzyılın başlarından itibaren istatistiksel programların eksiksiz veri setleri üzerine kurulması kayıp verilerin öneminin çıkış noktasını oluşturmaktadır. Kayıp veriler ile ilgili ilk önemli araştırmalar Afifi ve Elashoff (1969), Hartley (1971) ve Rubin (1987) gibi araştırmacılar tarafından yapılmıştır. Fakat en önemli gelişmenin ise Little ve Rubin tarafından geliştirilen ‘Analysis with Missing Data’ ve Rubin’in hazırladığı ‘Multiple Imputation for Nonresponse in Surveys’ adlı kitapların yayınlanmasıdır. Bu gelişmeler ile kayıp verilerin yol açtığı problemler üzerinde durulmaya başlanmış ve kayıp veri ile başa çıkacak istatistiksel analizler geliştirilmiştir (Graham, 2009). Bu istatistiksel programlar geliştirilirken genellikle sürekli veriler üzerinden çalışma yapmak için düşünülerek geliştirilmiştir.

İki kategorili veri setleri bir özelliğin varlığını ve yokluğunu temsil eder ve bu veriler kesikli veri olarak adlandırılır. Bu veriler genellikle evet-hayır, doğru-yanlış vb. gibi seçenekler sunmaktadır. Başarı testlerinde kullanılan ‘çoktan seçmeli’ maddeler de puanlama biçimi olarak iki kategorili maddeler arasında yer almaktadır (Erkuş, 2003). İki kategorili diğer bir ifadeyle kesikli verilerin, ‘yanlış’ kabul edilmesi veya ‘uygulanmamış’ olarak veri setine kodlanması analizlerde yanlı veya hatalı sonuçlara yol açmaktadır. Bu iki yaklaşımda da hatalı ve yanlı kestirimlerin oluşmasına yol açmaktadır. Pratikte kayıp veri olmadan bir veri seti elde etmek oldukça güçtür (Hohensinn ve Kubinger, 2011).

Kayıp verilerin analize dahil edilmemesi için kayıp verilerin ihmal edilebilir olduğuna dair kanıt ile desteklenmesi gerekmektedir. Kayıp verilerin ihmal edilebilir olup olmaması kayıp verinin meydana gelme süreçleriyle ilgilidir. Bu süreçte kayıp veri örüntüsü ve mekanizması, veri setine uygun analizin belirlenmesi ve yorumlanmasında önem arz etmektedir (Little ve Rubin, 1987).

Veri setindeki kayıp veri mekanizmasının ihmal edilebilir olması araştırmacının işini kolaylaştırırken, ihmal edilebilir olmayan kayıp veri mekanizması için ise oldukça önemli ön bilgilere ve incelemeye ihtiyaç duyulmaktadır. İhmal edilemez kayıp veri içeren veri setleri hangi modelin daha uygun olduğuna dair yeterli bilgi vermez. Bu sebeple araştırma sonuçları seçilen modele karşı yoldukça duyarlı olacaktır (Schaffer, 1997; Allison, 2002).

Yapılan sınavlarda soyut yapıları ölçmek çok zor ve bu yapılar ancak dolaylı yollarla ölçülebilir. Bu yapıları ölçmek sadece tek bir yaklaşımla sınırlı değildir. Bu ölçümleri yapan alan test kuramları olarak tanımlanmaktadır. Test kuramlarının genel

amacı ölçülmek istenen yapıya ilişkin hataları en aza indirmek ve en doğru şekilde yapmaktır (Crocker ve Algina, 1986). Ölçülmek istenen yapının gerçek değeri ölçmeye katılan çeşitli hatalardan dolayı, gerçek değere tam olarak elde edilemez; elde edilen ölçme sonuçlarına göre kestirilmeye çalışılır. Bu sonuçları elde etmek için istatistiksel kuramlar geliştirilmiştir (Baykul,2000). Ölçme çalışmalarının tarihi geçmişi incelendiğinde iki temel kurama rastlanmaktadır. Bu kuramlar kullanıma başladığı yıllar göz önüne alındığında ilk olarak “Klasik Test Kuramı” olduğu gözlenmektedir. Bu kuramda bazı sınırlılıkların olması sebebiyle 20. Yüzyıl ortalarında yeni bir kuram ortaya çıkmıştır. Bu karam “Madde Tepki Kuramı” veya “Örtük Özellikler Kuramı” olarak isimlendirilmektedir (Crocker ve Algina, 1986). Eğitim ile ilgili çalışmalarda araştırmacılar genellikle KTK ile çalışma yapmaktadır. Bunun en büyük sebebi KTK'nın zayıf varsayımlarının ve parametreleri kestirme kolaylığının bulunmasıdır (Hambleton ve Jones, 1993; Kelecioğlu, 2001).

1.1.1. Problem Cümlesi

Bu araştırmada Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE-2016) sınavında, Matematik başarı testindeki A kitapçığının çoktan seçmeli maddelerinin psikometrik özelliklerinin farklı kayıp veri yöntemlerine göre parametre kestirimlerinin karşılaştırılmıştır.

1.1.2. Alt Problemler

Verilen problem cümlesi göz önünde bulundurularak yapılan araştırmanın, alt problemleri aşağıda ifade edilmiştir.

1. ABİDE-2016 matematik testindeki kayıp veriler için nasıl bir örüntü ve mekanizma söz konusudur?
2. Liste bazında silme, Hotdeck atama, çoklu değer atama ve stokastik regresyon; kayıp veri teknikleri ile elde edilen madde parametreleri nelerdir?
3. Liste bazında silme, Hotdeck atama, çoklu değer atama ve stokastik regresyon; kayıp veri teknikleri ile elde edilen test parametreleri nelerdir?
4. Liste bazında silme, Hotdeck atama, çoklu değer atama ve stokastik regresyon ile elde edilen faktör analizi sonuçları nasıldır?
5. Liste bazında silme, Hotdeck atama, çoklu değer atama ve stokastik regresyon elde edilen güvenilirlik değerleri nelerdir?

1.2 Araştırmanın Amacı

Yapılan bu çalışmanın amacı ABİDE-2016 matematik testinin psikometrik özelliklerinin farklı kayıp veri baş etme teknikleri ile karşılaştırılmasıdır. Bu amaçla ABİDE-2016 A kitapçığı matematik testinin çoktan seçmeli maddelerinin kayıp veri yöntemlerine göre parametre kestirimleri gerçekleştirilmiştir.

1.3 Araştırmanın Önemi

Büyük bir hız ile gelişen dünyamızda eğitim ile ülkelerin gelişmişlikleri arasındaki bağlantı, eğitime büyük önem vermeyi beraberinde getirmektedir. Eğitime verilen bu önemle birlikte eğitimde ölçme ve değerlendirme çalışmaları büyük önem arz etmektedir (MEB,2015).

Eğitime verilen bu önem beraberinde geniş ölçekli sınavlarında yapılmasını beraberinde getirmektedir. Geniş katılımlı bir sınav ile her ülke kendi eğitim sisteminin eksik yönlerini ve diğer ülkelerden farklı olan yönlerini belirleme imkânına sahip olmaktadır. Bu geniş katılımlı sınavlar ile ilgili yapılan araştırmalar sonunda elde edilen veriler göz önüne alınarak eğitim sisteminde reformlar yapılmasına olanak sağlamaktadır. Literatür incelendiğinde PISA ve TIMSS gibi uluslararası sınavlar ile ilgili çok sayıda araştırma ve çalışma olduğu görülmektedir.

Uluslararası ve ulusal sınavlarda kayıp veriler ile ilgili yapılan çalışmalar incelendiğinde analizlerin yanlış olduğu bilinmektedir. Ölçme ve değerlendirmenin doğasında olan bu durum, giderilmesi gereken bir sorun teşkil etmektedir. Ancak araştırmacılar uzun yıllar boyunca bu sorunu dikkate almayıp analiz dışında bırakmışlardır. Bu sorun zamanla araştırmacıların çeşitli yöntem geliştirmesine yönlendirmiştir.

Veri setlerindeki kayıp veri miktarı, veri setinin büyüklüğü ve verilerin kesikli veya sürekli olması bu yöntemlerin seçilmesinde dikkat edilmesi gereken özelliklerdir. Yapılan çalışmaların analiz süreçlerinde kayıp verilerin dikkate alınmasının elde edilecek bilgi ve kanıtlar açısından gereklilik olduğu ve daha az yanlışlık ve sistematik hata oluşmasının önüne geçtiği görülmektedir. Bu nedenle ölçme sonucunda elde edilen parametrelerin doğru bir şekilde yorumlanabilmesi için kayıp veri sorununun çözümlenmesine yönelik çalışma gerekmektedir.

Özellikle iki kategorili puanlanan testlere yönelik kestirimler diğer sürekli verilere yönelik kestirimlere göre daha karmaşık bir yapıya sahiptir. Kayıp veri yöntemlerine yönelik çalışmalar gerçek verilerden çok yapay veriler üzerinden yapılmaktadır. Bu çalışmada ise gerçek veriler üzerinden Akademik Becerilerin izlenmesi ve Değerlendirilmesi (ABİDE) programının kayıp veri yöntemlerine göre elde edilecek parametrelerin büyük ölçekli testlerde kullanılan kayıp veri kullanma yöntemlerine katkısı olacağı düşünülmektedir. Dolayısıyla Türkiye’de yapılan kapsamlı çalışma olan ABİDE programı verilerinin analizinin yorumlanması ve yeni çözümler üretilmesi eğitim programının geliştirilmesine yönelik alan yazına katkı sağlayacağı düşünülmektedir.

1.4 Sayıtlar

Öğrencilerin testlerden elde ettikleri toplam puanların gerçek başarılarını yansıttığı varsayılmıştır.

1.5 Araştırmanın Sınırlılıkları

Bu araştırmadaki sınırlılıklar;

1. ABİDE-2016 uygulamasında yer alan 8. Sınıf matematik testi A kitapçığının çoktan seçmeli 9 maddesi ile sınırlıdır.
2. Bu çalışmada temel alınan Klasik Test Kuramı ile sınırlıdır.
3. Çalışmada kullanılan dört farklı kayıp veri yöntemleri ile sınırlıdır.

İKİNCİ BÖLÜM

KAVRAMSAL ÇERÇEVE

Kavramsal çerçeve başlığı altında; ABİDE programı ve ölçme hakkında bilgi verildikten sonra, Klasik Test Kuramı ve kayıp veri ile kayıp veri yöntemleri hakkında bilgilere yer verilmiştir.

2.1. Ölçme ve Değerlendirme

2.1.1. Ölçme

Günümüzde ölçme ifadesi ile çoğu yerde karşılaşılmaktadır. Örnek vermek gerekirse “oğlunuz TEOG’den kaç puan aldı?” veya “çocuğun YKS sınavından kaç puan aldı?” “Meteoroloji verilerine göre yarın hava sıcaklığı 38 derece olacak ve mevsim normallerinin üstünde olması bekleniyor.”, “Boyunuz kaç metre kilonuz ne kadar ?” vb. Bu ifadeler ölçme ile ilgili olduğu için ölçmenin bir tanımının olması gereklidir. Bu tanımları yapan araştırmacılardan , Chumney’e (2012) göre ölçme, veri toplamının ve uygulamaya dökmenin arkasındaki en önemli sebeptir. Küçükahmet’e (2004) göre ise ölçme, gözlenen niteliklerin (değişkenlerin), gözlem sonuçlarının sayı veya sembollerle ifade edilmesine denir. Tekin (2017), ise daha geniş bir ifadeyle herhangi bir kişi veya nesnenin veya nesnelerin herhangi bir özelliğinin olup olmadığını, eğer bir özelliğe sahip ise sahip olduğu miktarın gözlenip, gözlenen sonuçlarının sembollerle belirtilmesidir. Büyüköztürk (2015), ise ölçmenin, gözlemlenen bir olaya belli bir standarta göre değer verme, sayısallaştırma işlemi olarak ifade etmektedir. Ölçme, bazı fiziksel nesnelerin niteliklerinin veya boyutlarının belirlendiği süreci ifade eder (Kizlik, 2012). Yapılan bu tanımlardan yola çıkarak bir öğrencinin TEOG veya YKS’den aldığı puanın, bir kişinin boy uzunluğunu kaç metre ve kaç kilogram olduğunu, ne gibi düşüncelere ve kişilik özelliklerine sahip olduğunu öğrenmek için sorulan sorular ölçme işlemine örnektir. Ölçmeye konu olan niteliklerden bazılarının (uzunluk, kütle vb.) doğrudan gözlenebilirliği yani ölçülebilir bir nitelik taşıdığı, bazıları ise (yetenek, başarı, tutum, ilgi, kişilik vb.) doğrudan değil dolaylı yollarla ölçülebileceği verilen örneklerde anlaşılmaktadır (Bahar ve diğ. 2012). Daha net anlamda söylersek ölçme nitel verileri sayısal verilerle nicel verilere dönüştürmedir (Tan ve Erdoğan, 2004). Ölçme; bilgi, beceri ve tutum üzerine deneysel verilerin belgelenmesi ve kullanılmasını içeren sistematik bir süreçtir. Eğitimde ölçme yapılmasının temel amacı öğrencinin öğrenme kapasitesini geliştirmeye çalışmaktır (Chumney, 2012).

Ölçme aslında ölçülmek istenen özelliğin gözlemlenmesidir. Daha kesin, daha nesnel ve daha güvenilir bir yol olması ölçmenin diğer gözlem tekniklerinden üstün olan tarafını ve ölçme ile elde edilen verilerin ya da sembollerin sayı türünden ifade edilebilmesine bağlıdır (Turgut ve Baykul, 2012). Eğitimdeki ölçmeler, değerlendirme verisine kaynak oluşturur. Ölçme, insanların günlük hayatında çok önemli bir yeri olan kavramlardandır. Günlük yaşantımızda yaptığımız davranışları farkında olmadan ölçme sonucuna göre değerlendiririz. Bir şeyin büyüklüğü, miktarı ve ağırlığı

hakkında karar verirken ölçmeye yine ihtiyaç duyarız. Bu durum hayatımızı ciddi manada kolaylaştırmaktadır.

2.1.2. Eğitimde Ölçme

Ölçme kavramı, öğrencilerin gelişimlerinin ve hazır bulunuşluk düzeylerinin ele alınmasına, öğrenmedeki güçlüklerin saptanmasına, öğrencilerdeki güçlü ve zayıf yönlerin belirlenmesine ve belli bir program yeterliliğinin saptanmasına katkı göstermektedir (Yaşar, 2014). Böyle bir perspektiften bakıldığında ise ölçme “ne kadar?” sorusu dikkat çekmektedir (Gronlund ve Linn, 1990). Bunu yanı sıra, ölçme bir karar verme işi değil karar vermeyi sağlayıcı değişkenlerin hakkındaki bilgilere ulaşılmasıdır. Geneline bakıldığı zaman ölçme, özellikler ve değişkenler hakkında ulaşılmakta olan sonuçların sayı ve sembollere karşılık ifade edilmesidir (Turgut ve Baykul, 2012).

Eğitimde ölçme kavramının en etkin rolü karar verme adımında sağladığı destektir. Ölçme işlemi kapsamında süreç boyu ölçülmek istenen özelliklerin saptanması ile başlamaktadır. Devam eden aşamalarda ise ölçülecek olan sayılar ve sembollere karar verilerek nitelikler ve nicelikler ortaya konulur (Tan ve Erdoğan, 2004). Gözlenen ile gözlenemeyen arasındaki ilişkiler, bir başka ifade ile ölçülebilen değişkenlerden gözlenemeyen değişkenlerin elde edilmesine dayanan yaklaşımlar ölçmenin temel konularını oluşturmaktadır. Bu yaklaşımları genel olarak iki sınıfta ele alınabilir. Bunlar sırasıyla; doğrusal (linear) ve doğrusal olmayan (non-linear) bağıntılar şeklindedir. Doğrusal bağıntılardan en yaygın olarak kullanılan yaklaşım klasik test kuramı (classical testing theory), doğrusal olmayan yaklaşımda ise örtük özellikler (latent trait theory) kuramıdır

2.2. Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE)

Eğitim-öğretim sistemimizde öğretim programının hedefi öğrencilerin okul hayatında öğrendiklerinden ziyade öğrendiği bilgileri gerçek hayatta neler yapabildikleri ne kadar aktarabildikleri üzerinde yoğunlaşmaktadır. Bu sebeple öğrencilerin öğrendiklerini günlük hayata ne kadar aktarabildiklerini belirlemek için üst düzey zihinsel becerilerin ölçülmesini gerektirmektedir. Çünkü bu güne kadar ülkemizde yapılan sınavların öğrencilerin okul hayatında öğrendiklerini günlük hayata aktarmasını belirlemede eksik ve yetersiz kalmıştır. Tüm bu düşünceler göz önüne alınarak 2014 yılında eğitim-öğretim sistemimizin değerlendirilmesi için bir plan

yapılmıştır. Bu plan çerçevesinde üst düzey zihinsel ve bilişsel becerileri ölçmek ve bu konudaki eksiği kapatmak için 2016 yılında Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) çalışması adı altında sınav geliştirilmiştir (MEB, 2017c). Öğrenilen üst düzey zihinsel becerilerin günlük hayata aktarılması ve günlük hayat ile ilişkilendirilmesi için uygun eğitim-öğretim programlarının olması gerekmektedir. Buradan hareketle kalıcı ve nitelikli bir eğitim-öğretimin olması için üst düzey becerileri ölçmeye, günlük yaşamla ilişkilendiren ve odaklanan uygun bir ölçme ve değerlendirme çalışması gerektiğini savunmaktadır (Demirtaşlı, 2010).

ABİDE programı ders içerik ve kazanımlarının dikkate alması yönü ile TIMSS (Uluslararası Matematik ve Fen Eğilimleri Araştırması), yetenek ve becerilerin ölçülmesi yönü ile PISA (Uluslararası Öğrenci Değerlendirme Programı) ile benzerlik gösterdiği gözlenmektedir (MEB, 2017c). Bu sınavlar genel hatları ile konuların alanları ve öğrencilerin güdü, tutum ve kendilerine yönelik görüşleri; öğretmenlerin nitelikleri, öğrenme şekilleri, okul ortamlarında ve aileleri ile ilgili anketlerin uygulanması ile yapılmaktadır. Böylece öğrencilerin daha iyi tanınması adına onlardaki öğrenme isteklerinin ve derslerindeki performansların öğrenme ortamları ile ilgili tercihlerin daha net şekilde anlaşılması sağlanır. Buna ilaveten sınavlardan elde edilmekte olan ölçme soruları ile ülkeler karşılaştırılarak, eğitim politikalarında revizyonlara gidilebilir (MEB, 2019). Buradan anlaşılacağı üzere ABİDE sınavı hazırlanırken uluslararası düzeydeki geniş ölçekli testler üzerinde çalışma yapıldığı ve bu çalışmaların ABİDE sınavının meydana getirilmesine katkı sağladığı gözlenmektedir.

Bu çalışma ile öğrenciden istenileni gerçekçi, karşılaştırılabilir ve uygun bir ölçüt (PISA ve TIMSS ile karşılaştırılıp) ile değerlendirip uluslararası geniş ölçekli testlerden elde edilebilecek tecrübe ile ulusal ve uluslararası alanda eğitim-öğretimde yeni reformist politikaların üretilmesine katkı yapacaktır. Abide programı ayrıca uluslararası alanda yapılan geniş ölçekli testler öğrencilerimizin başarısına da olumlu yönde etki ederek eğitim-öğretim sistemimizi geliştirilmesine de katkı yapacaktır.

Türkiye’de 2016’dan beri üç senede bir uygulanmakta olan ABİDE projesi sayesinde sekizinci sınıf öğrencilerinde zihinsel becerilerin ölçülebilmesi ve bu becerilere yönelik öğretmen, aile ve okul özelliklerinin saptanması amaçlanmaktadır. PISA sınavındaki genel amaçlara benzer olarak Türkiye kültürüne özgü tamamlayıcı

bir rol oynayan ABİDE projesinin ilk çalışmaları İlköğretim Öğrencilerinin Başarılarının Belirlenmesi (ÖBBS) sınavıdır. Türkiye'deki öğrencilerin okul öğrenmelerini izleyebilmek amacıyla uygulanmakta olan bu sınavda öğrencilerin anketleri, öğretmenlerin anketleri ve başarı testi olacak şekilde üç ölçme aracı mevcuttur. ÖBBS ilköğretim düzeyinde 2002'den bu yana uygulanmaktadır. ÖBBS genel hatlarıyla benzer biçimde geliştirilmiş olan ABİDE sınavı kapsamında öğrenciler için matematik, fen, Türkçe ve sosyal bilimler derslerinin çoktan seçmeli ve açık uçlu maddelerinin yer aldığı 20 maddelik testlerdir. Buna ilave olarak öğrenci, öğretmen ve okul anketlerinin uygulanması ile öğrencilerin bilişsel düzeylerinin yanı sıra duyuşsal, okul, öğretmen ve ailevi özelliklerinin de saptanması amaçlanmaktadır. 2016 ABİDE projesi kapsamında uygulanmakta olan test ve ölçekler ilk olarak 2015 senesinde pilot bir uygulama ile 5000 öğrenci, 300 öğretmen ve 26 okul yöneticisi ile yapılmıştır. Ölçek ve testlerdeki ön uygulamalar için uzman görüşleri geçerlik ve güvenilirlik çalışmalarının yapılması ile 2016 senesinde ilk test ve ölçekler kullanıma başlanmıştır. ABİDE projesi bünyesinde test ve ölçekler 81 il bazında neredeyse 35000 öğrenciye uygulanarak; cinsiyet, kardeş sayısı, evdeki kitap sayısı, aile gelir düzeyi, derslere ilişkin devamsızlık, sosyal ve rehberlik faaliyetlerine katılma, okul içi ve okul dışı kurslara katılma vb. değişkenlere ek olarak ayrıca 16 ayrı ölçek uygulanmıştır. Tüm bunlar, okullara yönelik olan tutumlar, akran baskısı, ailelerin ilgi düzeyleri, aile baskıları, ilgili derslere olan sevgi ve değerler gibi ölçeklerden meydana gelmektedir (MEB, 2017c).

ABİDE programı PISA ve TIMSS'ten farklı olarak Türkçe, Matematik, Fen Bilimleri ve Sosyal Bilgiler olmak üzere dört farklı dersi göz önünde bulundurarak hazırlanmıştır (MEB,2017c). PISA Matematik okuryazarlığı, Fen okuryazarlığı ve Okuma becerileri olmak üzere üç alandan, TIMSS ise Fen ve Matematik olmak üzere iki alanda değerlendirme yapmaktadır. ABİDE programında dört alanda birden değerlendirme yapması, PISA ve TIMSS gibi sınavlara nazaran daha geniş öğretim alanını içine alıp ölçmektedir.

2.2.1. ABİDE Araştırmasının Genel Amacı

ABİDE Araştırmasının genel amacı “ortaokul sekizinci sınıf öğrencilerindeki bilişsel becerilerin saptanması ve öğrencilerin başarılar ile ilişkili öğrenci, öğretmen ve okul özelliklerinin tespit edilmesi” şeklinde belirlenmiş bulunmaktadır. Bu genel

amaç doğrultusunda hedeflenmekte olan alt amaçlar ise şu şekilde sunulmuştur (MEB, 2017c):

1.Öğrencilerin sekizinci sınıf seviyelerinde Fen Bilimleri, Türkçe, Matematik ve Sosyal Bilgiler alanlarında bilişsel becerilere hangi düzeyde sahip olduklarını belirlemek,

2.Öğrencilerdeki başarılarla etki etmekte olan duyuşsal, aile ve okul özelliklerinin saptanması,

3.Öğrencilerin puanlarına karşılık olarak gelen yeterlik seviyelerinin olduđu ve puanların anlam kazandıkları ölçme ve değerlendirme sisteminin geliştirilmesi,

4.Eğitim sistemlerinin geliştirilmesi adına bunları takip etmek için periyodik olarak ABİDE'yi uygulamak ve takip etmek.

2.2.2. ABİDE-2016 Araştırmasının Yöntemi

2016 senesinde düzenlenmekte olan çalışma kapsamında toplamda 16118 ortaokul ve 48091 şube dâhil edilmiştir. Çalışmalara örneklem olarak il başında yaklaşık olarak 400 öğrencinin katılmasının yeterli olacağı ön görülmüştür. Belirlenmiş olan sayılara şubeler aracılığı ile ulaşılırken 400 adet öğrencinin her bir il adına kaçar adet şubeye denk gelecekleri hesaplanmıştır. Evrenin temsili için veri toplanırken büyükşehirlerdeki şubelerin sayısına fazlaca yer verilmiştir. İl başına seçilmiş olan örneklemelerin başlangıcı kent-kır, ardından devlet-özel olarak sınıflandırılmıştır. Daha sonrasında okullar ikili-normal öğretim ve her ikisine de sahip okullar olarak kategorize edilmiştir. Son olarak ise okullar genel, yatılı ve imam hatip olarak sınıflandırılmıştır.

Çalışma kapsamında soruların yazarlarına açık uçlu olarak ve çoktan seçmeli olarak soruları yazmalarına yönelik eğitimler verilmiştir. Eğitimlerin ardından farklı soru türlerinin kullanımı ile dört alanda başarı testleri tasarlanmıştır. Soruların her biri ilgili alanların öğretim programlarından faydalanılarak meydana getirilen becerilerin göz önüne alınması ile tasarlanmıştır. Akademisyenlerin ve alan uzmanlarının katılımının kaynaklık ettiği soru yazımına; ölçme ve değerlendirme uzmanları ile dil uzmanları da dâhil edilmiştir .

ABİDE projesindeki ilk pilot uygulama 3 Haziran 2015 tarihli olarak Ankara'da yapılmış olup evren ve örneklem olarak 26 adet ortaokuldan seçilen 5000

öğrenci, 300 öğretmen ve 26 okul yöneticisinin katılımıyla gerçekleştirilmiştir. Pilot uygulamalar üzerinden seçilen öğrencilere yönetilmiş olan açık uçlu sorular sayesinde Milli Eğitim Bakanlığı tarafınca geliştirilmiş olan değerlendirme yazılımları sayesinde eğitim almış olan öğretmenlerin değerlendirmeleri yapılmaktadır. Cevapların her biri bağımsız iki değerlendirici kişi tarafından puanlandırılmıştır. Puanların aralarında farkın yer almaması halinde değerlendirme eğitimlerini vermekte olan üst değerlendiricilerin incelemeleri sonucu sonuçlar belirlenmiştir.

Pilot uygulamalardan elde edilen sonuçlara göre başarı testleri ve anketlerin gözden geçirilmesinin ardından ABİDE projesinin asıl uygulanmaya başlanması 2016 yılı Nisan-Mayıs ayları süresince seçilmiş olan örneklemdaki 81 ilden yapılmıştır. Asıl uygulamada bütün alanlar için 51 adet soruya yer verilmiştir ve bu sorular üzerinden 27 tanesinin pilot uygulaması yapılmıştır. Kalan 24 soru madde havuzunun oluşturulması için geliştirilerek öğrencilerin başarılarının yorumlanabilmesi için kullanılmıştır. Fen bilimleri alanında öğrencilere 13 tanesi çoktan seçmeli olan 14 tanesi açık uçlu olan toplam 27 soru yöneltilmiştir.

Görev yapılmakta olan bölge ve cinsiyet değişkenleri dikkate alındığında her alandan 15 adet öğretmen seçilmiştir. Derslerin her biri için üç adet masa oluşturulmuştur ve her masaya beş adet öğretmen verilmiştir. ABİDE süreciyle ilgili olarak önceki yapılmış olan çalışmalara katılan, süreci tanıyan öğretmenler de masa lideri olarak seçilmiştir. İlk olarak bütün katılımcılar için ABİDE ve standart belirleme çalışmalarıyla ilgili tanıtıcı sunumlar yapılmıştır. Ardından tüm grup öğrencilerinin tanımları ile ilk taslak meydana getirilmiştir. Hedeflenen öğrenci tanımlamaları, yeterlilik düzeylerinden herhangi bir tanesinde yer alan bir bireye ait olması istenen bilgi ve becerilerden oluşmaktadır. Tanımlar meydana getirilirken “ABİDE Beceri Belirleme Çalıştayı” sürecinde ortaya konulanlar ve PISA ve TIMSS yeterlik tanımlamalarından faydalanılmıştır. Devam eden süreçlerde masa liderleri de eğitim görmüştür. Tüm dersler için grup liderleri ölçme ve değerlendirme uzmanlarından seçilmiştir. Sürece yönelik uygulamalı eğitimler vermeye başlanmıştır.

ABİDE kitapçığı kapsamında bulunan maddelerin cevaplandırılması, katılımcıların eş zamanlı olarak öğrencilerin gözünden cevaplandırması ile sağlanmıştır. Ardından katılımcılar kolaydan zora giderek sıralanan soru kitapçıkları ve madde haritaları ile ilgilenmiş ve bunlarla ilgili cevapları açıkça vermeleri

istenmiştir. Tartışmalar ışığında masa liderleri “hedeflenen öğrenci tanımlarını dikkate alıp” katılımcıların altı-temel, temel-orta, orta-orta üstü, orta üstü-ileri şeklinde dört işaretleme yapmalarını istemiştir. İşaretlemelerin ardından katılımcıların cevapları tartışılmıştır. Son turda ise yapılmış olan işaretlemeler ortalamaları alınarak kesme noktası ile belirlenmiş ve kesme noktaları, yeterlik düzeylerinde bulunan maddelerin dikkate alınması ile yeterlik düzeyi tanımlamalarının son şekilleri verilmiştir. (MEB, 2017c).

2.3. Kayıp Veri Mekanizmaları

Eğitim-Öğretimde tüm dünyada uluslararası, ulusal, bölgesel, şehir ve okul bazında sınavlar olmaktadır. Okulda uygulanan testler, ulusal çapta ve uluslararası düzeyde yapılan geniş ölçekli testler öğrenciler, öğretmenler ve politikacılar için önem arz etmektedir. Çünkü eğitim öğretimden çıkan çıktılar ülkelerin politikalarını yönlendirmektedir. Bu nedenle yapılan bu sınavların güvenilir ve geçerli olması büyük önem arz etmektedir. Ölçme sonuçlarının geçerliliğini ve güvenilirliğini ölçmek için birçok çalışma vardır. Geçerlilik ve güvenirliliğin geliştirilmesi için bazı kanıtlar elde etme amacıyla farklı istatistik tekniklerine başvurulmaktadır. Bu istatistik tekniklerin yapı taşı eksiksiz veri seti üzerine kuruludur. Analizi yapan uzmanlar ne kadar dikkatli ve planlıda olsa, uygulama sonucunda kayıp verilerin olması, ölçme sonucu ve değerlendirmesi için bir sorun oluşturmaktadır. Bu sorun ise elde edilen bulguların yanlış ve eksik yorumlanmasını veya yorumları sınırlamaktadır.

Ölçme aracının genel amacı maddeye verilen cevaplardan yani gözlenen değişkenlerden, gözlenemeyen yani gizil değişkenler ile ilgili çıkarımlar yapmaktır. Yapılan bu ölçümlerde ölçme aracının uygulandığı kişilerin, cevaplaması gereken bazı maddeleri atlaması yani maddeyle ilgili gözlenen bir verinin elde edilememesi, ölçülmesi amaçlanan gizil değişken ile ilgili bilgi elde edilmesini zorlaştırmaktadır (Hohensinn ve Kubinger, 2011).

Eğitimde ve psikolojide ölçülmek istenen değişkenler doğrudan gözlenemeyen gizil değişkenlerdir. Doğrudan gözlenemeyen bu değişkenler gözlenebilen başka değişkenlerle ölçülmektedir. Ölçülmek istenen gizil değişkene bireylerin gözlenen değişkene verdikleri tepkiden yola çıkarak ulaşılmaktadır. Gizil değişkenlerin ölçülmesi için geliştirilen ölçme araçlarının titiz bir şekilde hazırlanması, söz konusu ölçme aracının geçerliliği ve güvenirliliğini sağlayacaktır (Akbaş, 2014).

Bireylerin testteki maddeye cevap verme olasılıklarının farklı olması, yanlılıktan mı yoksa ölçülen özelliğin cevaplayıcılardan farklı olmalarından mı kaynaklandığına dikkat edilmelidir (Camili ve Shepard, 1994; Clauser ve Mazor, 1998). Bir testin geçerlilik ve güvenilirliğini tehdit edecek en önemli unsurlardan biri, test sonuçlarının hatalı veri vermesine sebep olacak olan faktörlerden biri olan kayıp verilerdir. (Garrett, 2009).

Ölçme yapılan çalışmalarda nicel veriler genellikle eksiksiz toplanmamaktadır. Bu kayıpların sebepleri sorulan soruları cevaplamak istememesi, reddetmesi ve unutulması gibi sebepler olabilir. Katılımcıların ölçme aracını ve soruyu tam olarak anlamaması, soruları ciddiye almaması, yorgun ve hasta olmaları gibi nedenler kayıp veri olmasını etkilerken; veri toplarken oluşan hataların yanı sıra veri girişi sırasında da yapılan hatalar kayıp veri oluşmasını etkiler. Araştırmacılar yapılan ölçme de oluşan kayıp veriler hakkında dikkatli davranmalıdır. Kayıp veri analiz etme yöntemlerinde eldeki verilere göre bazı varsayımları gerektirmektedir. Bu varsayımları dikkatli kullanmadıklarında kayıp verilerde, yanlı ve yanıltıcı sonuçlarla karşılaşabilirler. Verilerin dağılımı ve kayıp olma nedenleri, uygulanan kayıp veri teknikleri için en önemli kriterdir (Pigott, 2001; Tabachnick ve Fidell, 2001).

Ölçme sonucunu analiz etmek için istatistiksel yöntemler geliştirilmiş ve genellikle bu yöntemler kullanılmaktadır. Geliştirilen bu yöntemlerdeki veri setlerinde satırlar gözlemler olarak ifade edilen kısımlardan, sütunlar ise ölçülen değişkenlerden oluşan matrislerden oluşmaktadır. Bu matrislerdeki girdiler; yaş, boy, kilo ve gelir düzeyi gibi sürekli değişkenlerden ve cinsiyet ve eğitim durumu gibi süreksiz değişkenlerden oluşmaktadır. Eldeki verilerde bulunan matrislerde boş hücrelerin olması bazı verilerin eksik olduğu anlamını taşımakta bu ise yapılan analizi olumsuz etkilemektedir (Little ve Rubin, 2002). Yapılan araştırmalar incelendiğinde kayıp verilerin, elde edilen sonuçların güvenilirliğini, geçerliliğini, standart sapma, ortalama gibi betimsel istatistikleri de olumsuz yönde etkilediği görülmüştür. (Little ve Rubin, 2002; Hohensinn ve Kubinger, 2011).

Ölçme sonucunda elde edilen veri setindeki verilerin herhangi bir örüntü sergileyip sergilemediğini anlamak için kayıp veri kullanma teknikleri kullanmak zorundadır. Yapılan analizler genellikle eksiksiz veriler üzerine yapılmakta olduğundan analiz sırasında, kayıp verinin olduğu satırların veri setinden çıkarılması, genellikle tercih

edilen bir yöntemdir (Akbaş, 2014). Genellikle istatistik programları veri setinde eksik verilerin olduğu satırın analiz dışında bırakmak için satırları sildiği, kayıp veriye sahip bireylerin oranının (%5 ve daha az) olduğu durumlarda savunulabilecek bir çözüm olduğunu ifade etmektedir (Schafer ve Graham, 2002).

Kayıp veri mekanizması, kayıp verilerin ve kayıp veri içeren değişkenlerin belirli özelliklerini anlamada ve incelemesinde ve inceleme sonucunda ortaya çıkacak bilgiler çerçevesinde kullanılacak kayıp veri ile başa çıkma yönteminin belirlenmesinde önemli bir yol göstericidir (Demir,2013). Kayıpların bulunduğu veri setlerinde, bu kayıpların hangi sebeple olduğunu tahmin edebilmek için hangi yöntemin kullanılacağı, kayıp verilerin niçin meydana geldiğini açıklayan kayıp veri mekanizmasına, başka bir deyişle kayıp verilerin rastlantısal dağılıp dağılmadığına bağlıdır (Little ve Rubin'nin, 1987).

Kayıp veriler için genel kabul gören sınıflama Rubin (1976) ve Little ve Rubin'nin (1987)'de yaptığı çalışmalara dayanmaktadır. Yapılan bu çalışmalara göre kayıp veriler ile veriler arasındaki ilişkiler göz önüne alınarak 1- Tümüyle Seçkisiz Kayıp (TSK/Missing Completely At Random-MCAR), 2-Seçkisi Kayıp(SK/Missing At Random-MAR), 3- Seçkisiz Olmayan Kayıp(SOK/Missing Not At Random-MNAR) şeklinde üç farklı mekanizmadan söz edilmektedir.

Little ve Rubin (2002), kayıp verinin meydana gelme sebeplerini üç kategoriye ayırmıştır.

- (i) Kayıp veri, gözlenen değerlerden bağımsız olarak meydana gelir ise, bu kayıp veri tamamen seçkisiz kayıptır.
- (ii) Bağımsız değişkenlerden biri diğer bağımsız değişken üzerinde etkisi var ise, ve mevcut değişken üzerinde bir etkisi bulunmuyor ise seçkisiz kayıptır.
- (iii) Değişkenler dağılımda karşılıklı olarak birbirini etkiliyor ise buradaki kayıplar seçkisiz olmayan kayıptır.

Tablo 1'de üç farklı kayıp veri mekanizması TSK, SK ve SOK mekanizmaları için örnek veri seti verilmektedir. Bu tabloda Ocak ve Şubat aylarında alınan tansiyon (mmHg) değerleri ile ölçümler bulunmaktadır (Schafer ve Graham, 2002).

Tablo 1. Üç Farklı Kayıp Veri Mekanizması için Örnek Veri Seti

Ocak ayı ölçümleri (X)	Şubat ayı ölçümleri (Y)			
	Eksiksiz	TSK	SK	SOK
78	74	74	-	-
94	113	-	-	-
97	85	85	-	-
101	119	-	-	-
103	106	106	-	-
105	138	-	-	-
106	148	-	-	148
109	96	-	-	-
109	78	-	-	-
111	129	-	-	-
112	100	-	-	-
116	102	-	-	-
118	118	-	-	-
125	88	-	-	-
126	123	-	-	-
128	155	-	-	155
130	101	101	-	-
131	131	-	-	-
132	149	-	-	149
133	150	-	-	150
134	124	124	-	-
136	140	-	-	-
137	122	122	-	-
145	155	-	155	155
146	134	-	134	-
151	113	-	113	-
153	112	-	112	-
160	169	-	169	169
169	148	148	148	148
176	137	-	137	-

Tablo 1’deki ilk sütun Ocak ayının eksiksiz ölçümleri, ikinci sütun ise Şubat ayının eksiksiz ölçümü yapılan tansiyon verilerini göstermektedir. Tablo 1’de üçüncü sütundaki değerler ikinci sütundaki değerlerden seçkisiz (rastgele) seçildiği görülmektedir. Buradaki ölçümler ne ocak ayındaki ölçümlerle nede şubat ayında ölçülen düşük ve yüksek tansiyonla ilgili olmadığı görülmektedir. Bu seçim bize TSK mekanizmasının bir örneğidir. TSK’dan söz edilebilmesi için veri setindeki herhangi bir verinin kayıp olma olasılığı ile bu değişkenin değerleri ile veri setindeki diğer değişkenlerden elde edilen veriler arasında bir ilişki olmaması ve birbirinden bağımsız olmasıdır. (Little ve Rubin,2002). TSK, Yani bir Y değişkeninin kendisiyle ve diğer

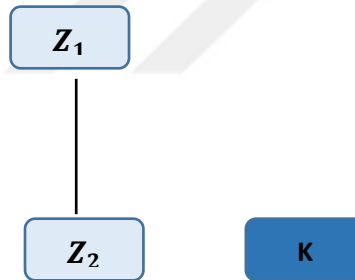
değişkenlerle ilişkisiz olduğunu ifade etmektedir. Bu varsayımın diğer tüm değişkenler tarafından karşılanması halinde veri setindeki gözlenen değerler, veri setinin üzerinde çalışılması için seçilmiş bir örnekleme olarak düşünülebilir (Allison, 2002; Enders, 2010).

X gözlenen, Y kayıp veri içeren değişken olmak üzere TSK mekanizmasının formüle edilmesi aşağıdaki gibidir;

$$P(Y_{kayıp} | X, Y) = P(Y_{kayıp})$$

Verilen bur formülde Y değişkeni aylık geliri, X değişkeni medeni durumu temsil etsin. TSK'ya göre aylık gelir değişkenindeki kayıp veri olma olasılığı ne aylık gelir ile ne de medeni durum ile arasında ilişki yoktur (Allison, 2003). TSK mekanizmasını daha net anlamak için;

$Z = Z_{ij}$ veri matrisini, $K = K_{ij}$ kayıp veri matrisini temsil etmek üzere, Z verildiğinde TSK için K'nın dağılımı aşağıda verilen Şekil 1'de gibi gösterilebilir. Bu şekilde iki değişkenli bir veri kümesini, arada verilen çizgiler ise istatistiksel ilişkiyi göstermektedir.



Şekil 1. Tamamen Seçkisiz Kayıp Mekanizmasına ilişkin Gösterim

Şekil 1'de birbiriyle ilişkili iki değişken olan Z_1 ve Z_2 bulunmaktadır. K, Z_2 değişkenindeki kayıp verileri ifade eden kayıp veri matrisini göstermektedir. K hem Z_1 hem de Z_2 ilişkisiz olduğu görülmektedir. Bu durum bize K ile belirlenen kayıp verilerin tamamen seçkisiz olduğunu göstermektedir.

Tablo 1'de dördüncü sütündeki seçilen değerler incelendiğinde Ocak ayında yapılan ölçümlerden hipertansiyon değeri 140'ın üzerinde olanlardan ($X > 140$) ölçümler alınmıştır. Bu sütündeki kayıp veriler incelendiğinde Şubat ayındaki ölçümlerdeki kayıpların Şubat ayı ölçümlerinden bağımsız olduğu, fakat yanlış olarak Ocak ayındaki ölçümlerle bağımlı olduğu görülmektedir. Veri kaybı olmasının Şubat

ayındaki ölçümlerin düşük veya yüksek olmasıyla ilişkili olmadığı fakat ocak ayındaki yüksek tansiyon ölçümleri ile yanlı olarak ilişkili olduğu görülmektedir. Bu ölçümler sonucunda elde edilen kayıp veriler bize SK mekanizmasının bir örneğini göstermektedir.

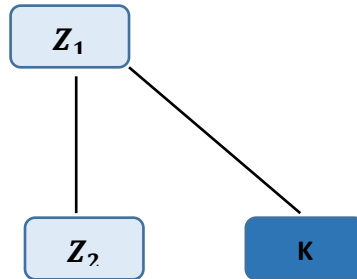
SK, verilerdeki kayıp olma olasılığının veri setindeki diğer değişkenlerle ilişkili olduğunu fakat kendi içinde ilişkili olmadığını ifade eder. Diğer bir ifadeyle veri setindeki diğer değişkenler kontrol altına alındığında, Y değişkenindeki kayıp olan verilerin Y değeri ile ilişkisi olmamasıdır. (Little ve Rubin, 1987).

X gözlenen, Y kayıp veri içeren değişken olmak üzere SK mekanizmasının formüle edilmesi aşağıdaki gibidir;

$$P(Y_{kayıp} | X, Y) = P(Y_{kayıp} | X)$$

Verilen bur formülde X ve Y iki tane test olsun. X sınavından başarısız olan bir birey Y sınavına girmek istememektedir. Buradan elde edilen sonuç Y sınavını yapma durumu X sınavına bağlıdır. Yani Y sınavının kendisi ile ilgisi yoktur. Bu nedenle Y sınavındaki kayıp veriler seçkisi kayıptır (Enders, 2010). Yani SK'ya göre kayıp veri olasılığı gözlenemeyen değişkenlerden değil de gözlenen değişkenlerden kaynaklanmaktadır. Kayıp veriye sahip bireylere ait değerler, gözlenen verilere sahip bireylerden daha düşük veya yüksek değerlere sahip ise kayıplar seçkisiz kayıp olarak değerlendirilemez. Buradan anlaşılacağı üzere SK varsayımının istatistiksel olarak test etmek mümkün değildir (Allison, 2003).

Z matrisi verildiğinde SK için K'nın dağılımı aşağıda Şekil 2'deki gibi gösterilebilir. Bu şekilde iki değişkenli bir veri kümesini, arada verilen çizgiler ise istatistiksel ilişkiyi göstermektedir.



Şekil 2. Seçkisiz Kayıp Mekanizmasına ilişkin Gösterim

Şekil 2'de K kayıp veri matrisi Z₂ değişkeninde meydana gelen kayıp verileri göstermektedir. Verilen şekil incelendiğinde K kayıp veri matrisinin Z₂ değişkeni ile

ilişkili olmadığı fakat Z_1 değişkeni ile ilişkili olduğu görülmektedir. TSK mekanizmasında kayıp verinin hiçbir değişken ile ilişkili olmadığı fakat SK mekanizmasında ise diğer değişkenlerle ilişkili olduğu görülmektedir.

Tablo 1’de beşinci sütundaki veriler incelendiğinde elde edilen ölçümlerin şubat ayındaki ölçümlerden 140’ın üzerinde olanlara ($Y > 140$) ait ölçümler olduğu görülmektedir. Bu sütundaki kayıp veriler incelendiğinde şubat ayındaki ölçülen kayıp verilerin ocak ayındaki ölçümlerden bağımsız olduğu fakat yanlı olarak şubat ayı ölçümlerine bağımlı olduğu görülmektedir. Veri kaybı olmasının ocak ayı ölçümlerinin düşük ya da yüksek olmasıyla ilişkili olmadığı fakat şubat ayı hipertansiyon alt sınırının üzerinde olanlarla yanlı olarak ilişkili olduğu görülmektedir. Diğer bir ifadeyle ölçümleri kaydeden bir kişi normal aralıktaki tansiyon değerlerini kaydetmemiş olabilir. Bu ise bize SOK mekanizmasının bir örneğini göstermektedir.

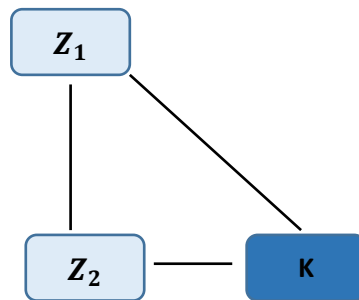
SOK, bir değişkendeki kayıp verinin kayıp olma olasılığının bu değişkenin kendisiyle ilişkili olduğunu ifade eder. Yani Y değişkenindeki kayıp verinin olma sebebi Y değişkeninin kendisiyle alakalıdır (Enders,2010).

X gözlenen, Y kayıp veri içeren değişken olmak üzere SOK mekanizmasının formüle edilmesi aşağıdaki gibidir;

$$P(Y_{kayıp} | X, Y) = P(Y_{kayıp} | Y)$$

Verilen bu formülde Y değişkeni suç işlemeyi, X değişkeni öğrenim durumunu temsil etsin. SOK’a göre suç işleme değişkenindeki kayıp veri olma olasılığı ile suç işleme arasında bir ilişki vardır (Enders,2010). Bu veriler ışığında SOK mekanizmasında kayıp veriler değişkenin kendisiyle ilişkili iken SK’da ise verilerin kayıp olma olasılığı diğer değişkenlerle ilgilidir.

Z matrisi verildiğinde SOK için K ’nın dağılımı aşağıda Şekil 3’deki gibi gösterilebilir.



Şekil 3. Seçkisiz Olmayan Kayıp Mekanizmasına ilişkin Gösterim

K, kayıp veri matrisindeki kayıplar Z_2 değişkenindeki kayıplardan oluşturulsun. Kayıp verinin Z_1 ve Z_2 değişkenleri ile ilişkili olduğu görülmektedir. Bu ilişki bize SOK mekanizmasını göstermektedir.

SK ve bu mekanizmanın özel bir hali olan TSK mekanizmasında kayıp veriler ihmal edilebilir, SOK ise ihmal edilemez kayıp veri olarak nitelendirilmektedir. Burada bahsedilen ihmal edilebilirlik önemsiz olma veya dikkate almama anlamına gelmemektedir (Schafer ve Graham,2002). Bu ifadede anlaşılacağı üzere ihmal edilip edilmemenin, kayıp verinin kendisinin değil kayıp veri mekanizması ile ilgili olmasıdır. Başka bir ifadeyle bir veri SK veya TSK özelliği gösteriyorsa bu durum herhangi bir modelle açıklamaya gerek duyulmaz. Ancak kayıp veri Seçkisiz Olmayan Kayıp (SOK) özelliği gösteriyor ise kayıplar bir değişkene bağlı fonksiyon ile gösterilebilir.

SOK mekanizmasındaki en büyük zorluk, sadece gözlenen verilere dayanarak yani kesin bir kanıt olmadan değişkenler hakkında varsayımda bulunmaktır. Örnek vermek gerekirse yapılan bir ankette, gelir düzeyi hakkında bilgi vermeyen bir kişinin niçin cevap vermediği hakkında kesin bir tahmin yapmak imkânsızdır (Molenberghs ve Kenward, 2007).

Graham (2012), kayıp veriye sebebiyet veren değişkenin ulaşılabilmesi ve analize dahil edilmesine göre kayıp veri mekanizmalarını Tablo 2’de özetlemektedir.

Tablo 2. Kayıp Veri ile İlişkili Değişkenlerin Analize Dahil Edilme Durumu

Kaybın Nedeni	Veri kaybına neden olan değişkenlerin analize dahil edilme durumu	
	Analize Dahil	Analize Dahil Değil
TSK	TSK	TSK
Ulaşıldı	SK	SOK
Ulaşılamadı	-	SOK

Tablo 2 incelendiğinde kayıp verilerin TSK mekanizmasına sahip olması durumunda analizlerde diğer değişkenlerin bulunmasına gerek olmadığı görülmektedir. Kayıp verilerin TSK olmadığı durumlarda ise veri kaybına neden olan değişkenin analize dahil edilmesi durumunda kayıpların SK olduğu, dahil edilmediğinde ise SOK mekanizmasında söz edilebilir. Veri kaybına sebep olan değişkenlerin bilinmediği veya ulaşılamadığı durumlarda ise; analize katılması

mümkün olmaması sebebiyle SOK mekanizmasından söz edilebilir. Başka bir ifadeyle kayıp verilerin TSK göstermesi durumunda araştırmacının elindeki veri seti amaçladığı veri setinin basit seçkisiz örnekleme olacaktır. Yani TSK mekanizmasının yanlılık içermeyeceği, fakat SK ve SOK mekanizmasına sahip kayıp verilerden elde edilen sonuçlar yanlı olabilir (Akbaş ve Koğar, 2020).

Allison (2009)'a göre kayıp verilere yönelik kullanılacak iyi bir yöntemin;

- (i) Veri kaybından kaynaklanacak yanlılığı en aza indirmesi,
- (ii) Elde edilecek olasılık değerleri, standart hata ve güven aralığı kestirimlerini titiz bir şekilde yapması,
- (iii) Gerekli ve elverişli bilgilere ulaşmayı maksimize etmesi beklenir.

2.3.1. Kayıp veri Sürecinde Rastgeleliğin Sorgulanması

Kayıp veri analizinde kullanılacak yöntemin belirlenmesi için rastgeleliğin aşağıdaki yöntemlerle araştırılması gerekir.

(i) Veri setindeki değişkenler, kayıp veri içeren ve içermeyen olmak üzere iki gruba ayrılır. Daha sonra bu iki grup ortalaması arasında anlamlı bir farkın olup olmadığına dair sonuçlara ulaşmak için “t-testi” yapılır. Anlamlı bir farkın olması bize kayıp verilerin rastgele dağıldığını gösterir (Alpar, 2003).

(ii) Veri setindeki değişkenler, kayıp veri içeren ve içermeyen olmak üzere iki gruba ayrılır. Daha sonra tam veriler 1, kayıp veriler 0 olarak kodlandıktan sonra değişkenler arasındaki korelasyon katsayısı hesaplanır. Elde edilen korelasyon değerleri değişken çiftleri arasında ne kadar ilişki olduğunu gösterir. Korelasyon katsayısının küçük olması rastgeleliliği gösterir (Alpar, 2003).

(iii) Son olarak TSK mekanizmasında rastgeleliği belirlemede kullanılan ki-kare testi olan ‘Little’s MCAR Test’ dir. Test sonucunda elde edilen kestirimler ($p < .05$) olduğu durumlarda mevcut olan veri yapısının TSK olmadığı sonucuna varılır (Little, 1998).

2.3.2. Kayıp veri ile baş etme yöntemleri

Yapılan çalışmalarda kayıp veri tekniklerinden bazı tekniklerin diğerlerine göre bazı üstünlüklerinin olmakla birlikte tamamen eksiksiz olarak tanımlanamayacağını ifade etmektedir. Yapılan analizde en iyi çözümün veri kaybının olmaması olacağını

fakat bu mümkün olmayacağından dolayı, kayıp veriyi en aza indirmek için bir çaba sarf edilmelidir. Kayıp veriler yapılan bütün istatistik analizlerde sorun yaratmaktadır. Kayıp veriler ile başa çıkmada uygun olmayan yöntemin kullanılması, parametre kestirimlerinde eksik veya yanlış bilgiler elde edilmesine sebep olacaktır. Bu sebeptir ki kayıp veriyi azaltmada kullanılan farklı kayıp veri ile baş etme yöntemleri kullanılmaktadır. Bu yöntemlerden uygun olanı seçmek verinin doğru yorumlanmasında büyük önem arz etmektedir (Allison, 2002).

Bu yöntemlerden silmeye dayalı en geleneksel olanları Liste Bazında Silme(Listwise Deletion-LW), Çift Bazında Silme (Pairwise Deletion) yöntemleridir. Atamaya dayalı yöntemlerden ise en sık kullanılan Ortalama Atama, Hot Deck Atama (HDA), Cold Deck Atama (CDA), Regresyon Atama (RA), Stokastik Regresyon ile Atama (SRA), Çoklu Değer Atama (ÇDA) ve Beklenti-Maksimizasyon (BM) yöntemleridir. Kayıp veriler analiz sonuçlarında yanlılığa sebep olacağından seçilen yöntemin yanlılığı en az seviyeye düşürmelidir. Kayıp veri yöntemlerinden iyi bir kayıp veri yönteminin, kayıp verilerin kısıtladığı varsayımlarına karşı gelişmiş olmalıdır. (Allison, 2002).

2.3.2.1. Silme Yöntemleri

En yaygın olarak kullanılan kayıp veri yöntemleridir. Veri silme yöntemleri liste bazında silme ve çift bazında silme olmak üzere iki şekilde elde edebiliriz.

Liste Bazında Silme

Bu yöntemde analizi yapılan değişkenlerde en az bir kayıp verinin bulunması durumunda, kayıp verinin bulunduğu satırın analiz dışına çıkarılması ile geriye kalan maddeler üzerinden analizin yapılmasına dayanır. Diğer bir ifadeyle analizin eksiksiz veriler üzerinden gerçekleştirilmesini sağlar. (Graham,2012). Fakat kayıpların analiz dışında bırakılması veri kaybına sebep olacağından standart hatanın artmasına ve güvenilirliğin azalmasına sebep olacaktır. Kayıp verilerin TSK mekanizmasına sahip olması geriye kalan eksiksiz veri setinin, orijinal olan veri setinin seçkisiz bir örnekleme olacağı için standart hata değerleri birbirine yakın sonuçlar verecektir. (Allison, 2003). Veri setinde analize dahil edilecek kayıpların TSK mekanizmasına sahip olması durumunda analiz yanlılık içermeyecektir. Fakat analize dahil edilen kayıpların SK ve SOK mekanizmasına sahip olduğu durumlarda liste bazında silme tekniği yanlılık gösterecektir. (Graham,2012).

Çift Bazında Silme

Çift bazında silme yöntemi, eldeki tüm verilerin kullanılmasına dayanan bu teknik, Erişilebilir hücre analizi olarak ta bilinen bu yöntem, tüm gözlem değerlerinden yola çıkarak bazı istatistik parametreleri önceden hesaplar. Bu yöntem korelasyon ve kovaryans matrisi ile açıklanmaktadır. Eldeki tam veriler kullanılarak korelasyon ve kovaryans parametreleri ile ortalama, standart sapma gibi istatistiksel kestirimleri içermektedir (Allison, 2003). Bu yöntemde eldeki veriler veri çiftlerine göre hesaplamadan silinir veya hesaplamaya katılır. Örneğin, bir veri setindeki katılımcının Y , X_1 , X_2 ve X_3 değişkenlerindeki verileri eksiksiz veri, X_4 ve X_5 değişkenlerindeki veriler ise eksik veriler olsun. Bu bağlamda korelasyon kestirimi yapılırken r_{YX_1} , r_{YX_2} , r_{YX_3} değişkenlerinin korelasyon değeri hesaplanırken, r_{YX_4} , r_{YX_5} değişkenlerinin korelasyon değeri hesaplanmaz (Howell, 2007). Bu yöntemin liste bazında silme yönteminden farkı, kayıp veri içeren verinin tamamının veri setinden çıkarmaktansa sadece kayıp veri içeren durumların çıkarılmasıdır (Cheema, 2012). Liste bazında silme ve çift bazında silme tekniğinin karşılaştırılmasına yönelik bir örnek Tablo 3’de verilmiştir (Akbaş ve Koğar, 2020).

Tablo 3. Liste Bazında ve Çift Bazında Silme Tekniklerinin Karşılaştırılması

Katılımcılar	Ölçümler		
	X_1	X_2	X_3
1	2	-	5
2	4	5	6
3	3	-	-
4	5	6	7
5	7	6	4
6	-	-	5
7	4	3	-
8	-	7	9

Tablo 3’de yer alan değişkenlere yönelik liste bazında korelasyon hesaplanacak ise, yapılacak parametre kestirimler herhangi bir değişkende kaybı olmayan 2,4 ve 5 numaralı katılımcılar üzerinden yapılacaktır. Çünkü liste bazında yapılan analizlerde veri kaybı olan katılımcılar analiz dışında bırakılmaktadır. Diğer yandan korelasyon

değerleri çift bazında yapılacak ise X_1 ile X_2 arasındaki korelasyon değerleri 2,4,5 ve 7 ; X_1 ile X_3 arasındaki korelasyon 1,2,4 ve 5 ; X_2 ile X_3 arasındaki korelasyon ise 2,4,5 ve 8 numaralı katılımcıların verileri ile hesaplanacaktır.

Buradan anlaşılacağı üzere analizler liste bazında silme ve çift bazında silme yöntemlerinde kayıp verilerin yerine herhangi bir değer atanmadan eksiksiz veriler üzerinden yapılmaktadır.

2.3.2.2. Atama Yöntemleri

Veri silme yöntemlerinin bilgi kaybına sebebiyet vermesi araştırmacıları kayıp verileri bulmada farklı kayıp veri atama yöntemlerini araştırmaya yönlendirmiştir. Bu nedenle araştırmacılar kayıp verileri belirlemede farklı atama yöntemlerine başvurmuşlardır.

Kayıp verilerde liste bazında ve çift bazında silme örneklem büyüklüğü üzerindeki olumsuz yöndeki etkisine karşı değer atama yöntemlerine başvurulabilir. Çünkü değer atama yöntemlerinde eksiksiz veri seti üzerinden kestirimler yapılmaktadır. Kayıp verilerin yerine atama yapılırken dikkat edilmesi gereken bazı noktalar vardır. Veri setindeki kayıp verilerin oranı, kayıp veri örüntüsü ve kayıp veri mekanizmasının durumuna göre hangi atama yönteminin kullanılmasına karar verilmelidir. Bu durumların dikkate alınmaması elde edilen kestirimlerin güvenli sonuçlar vermemesine sebep olacaktır (Enders, 2010; Schafer ve Graham, 2002). Bu çalışmada en sık kullanılan kayıp veri atama yöntemlerinden bahsedilmiştir.

Ortalama Atama

Bu yöntemde veri setindeki kayıp veriler yerine, gözlenen değerlerin ortalaması alınarak bir atama yapılmaktadır (Baraldi ve Enders, 2010; Little ve Rubin, 2002). Kayıp veriler ile ilgili elde bir bilgi yok ve kayıp veriler rastgele dağılım gösteriyor ise bu ortalama yöntem iyi sonuçlar vermektedir. Fakat kayıp veriler rastgele dağılıma sahip değil ise bu yöntem yanlış sonuçlar vermektedir. Kayıp veri içeren değişkenlerin hepsine aynı değerlerin atanması merkeze doğru bir yığılmaya sebep olduğundan, elde edilen varyans, gerçek varyanstan düşük çıkmaktadır (Little ve Rubin, 1987). Varyansın düşmesine paralel olarak korelasyon ve kovaryans değerlerinin de düşmesine aynı zamanda yanlış tahminlere yol açmaktadır. Veri setinde kayıp verinin

fazla olması yanlılığı arttırmaktadır. Bu sebeple ortalama atanmanın kullanılmaması gerektiği belirtilmektedir (Enders, 2010; Pigott, 2001).

Hot-Deck Atama

Veri setinde birbirine benzeyen fazla sayıda kayıp veri olması durumunda, benzer gözlemlerin çok olması aynı gözlemlerin tekrar kullanılmasının önüne geçecektir. Birbirine benzeyen az sayıda kayıp verinin olması durumunda ‘benzerlik’ kavramının net olarak anlaşılması açısından olumludur. Bu yöntemin avantajı kavramsal olarak kolay anlaşılması ve kayıp verileri tamamlanan değişkenin varyansını azaltmamasıdır. Diğer kayıp veri atama yöntemlerindeki gibi standart hatayı ciddiye almama eğiliminde (Enders, 2010; Little ve Rubin, 2002); Korelasyon ve regresyon katsayısı kestirimlerinde ise yanlılık göstermektedir (Schafer ve Graham, 2002).

Bu yöntemde eksik veriler veri setinde daha önce gözlenmiş veriler ile doldurulur. Yani bu yöntemin ana prensibi kayıt altına alınmış güncel verilerin atanmasıdır. Kayıp verilerin ataması yapılırken satırlar arası uzaklık olan k-en yakın komşu hesabı yapılmaktadır. Bu yöntemin uygulanması için aşağıdaki adımlar takip edilir ;

-Kayıp verileri içeren küme ve tamamlanmış veriler olmak üzere iki grup oluşturulur.

- X_i tamamlanmış veri matrisini, X_{ij} i. durumun j. değişkenini; Y_i tamamlanmamış veri matrisini, Y_{ij} i. durumun j. değişkenini belirtmektedir.

-Daha sonra bu kümeler üzerinden, kayıp veri içeren satır için Öklid uzaklığı (farklı uzaklık ölçüleri de dikkate alınabilir) (d) hesaplanır.

$$\text{Öklid}(d) = \sqrt{\sum_{j=1}^n (X_{ij} - Y_{kj})^2}$$

Öklid (d) : Öklid uzaklığı

X_i : Tamamlanmış veri küme matrisi

X_{ij} : i. durumun j. değişkeni

Y_i : Tamamlanmamış veri kümesi matrisi

Y_{ij} : i. durumun j. değişkenidir.

Cold Deck Atama

Bu atama yönteminde daha önce yapılmış güvenilir çalışmalardan ortalama ve ortalamaya yakın özellikteki sayısal kestirimler kayıp verilerin yerine atanır. Bu yöntemdeki veriler geçerliliği yüksek kaynaklardan alınan verilerdir. Dış kaynaklardan alınan veriler, eldeki verilerden elde edilen kestirimlerden daha geçerli olabileceğinden emin olmalıdır. Bu yöntemin ortalama atama ile benzer olumsuz yönleri bulunmaktadır (Alpar, 2003).

Regresyon Atama

Bu yöntemde veri setindeki kayıp veriler bağımlı değişken, diğer veriler ise bağımsız değişken kabul edilerek kayıp verilerin ataması yapılır. Kayıp verilerin ataması yapılırken tam veri seti üzerinden regresyon denklemi kurulur ve bu denklem üzerinden kayıp verilerin ataması yapılır. Regresyon ile yapılan atama araştırmacının yaptığı (Enders, 2010; Tabachnick ve Fidell, 2001). Bu yöntemde aslında yapılmak istenen elde olan veriler ile kayıp olan verileri elde etmektir.

Bu atama yöntemi kullanılırken dikkat edilmesi gereken noktalar vardır. Bağımlı değişkenin bağımsız değişken tarafından yeterince açıklanması ve değişkenler arasındaki ilişkinin belirli bir seviyede olması gerekir. Bu durumların sağlanmadığı zaman regresyon ataması yerine ortalama atamanın kullanılması daha uygundur (Alpar, 2003). Bu yöntemin olumlu yönlerinin yanında olumsuz yönleri de bulunmaktadır. Bu yönleri aşağıdaki gibi sıralanabilir.

- I. Bu yöntemde elde edilen kestirimler eldeki veriler üzerinden olacağı için değişkenler arasındaki ilişkiyi daha fazla güçlendirmektedir. Bu durum gerçek hayattaki değerlerden daha yakın değerler üreteceğini göstermektedir.
- II. Bu yöntemde atanan kayıp veriler eldeki kayıp verilere yakın değerler olacağı için varyans değeri azalacaktır.
- III. Bu yöntemde kayıp veriyi tahmin etmede kullanılan bağımsız değişkenler kayıp olan verileri yanlış olarak kestirirse, regresyon atamanın bu durumda ortalama atamadan farkı olmayacaktır (Tabachnick ve Fidell, 2013).

Stokastik Regresyon ile Atama

Kayıp veri yöntemlerinden olan bu yöntemin regresyon atamasından ayıran yönü, regresyonla belirlenen değerlere bir hata teriminin eklenmesi ile oluşmaktadır. Bu hatanın regresyon tekniğine eklenmesi ile regresyon değer atama yönteminde bir değişkenlik meydana getirmektedir. Bu yöntemin SK gibi durumlar da bile yansız kestirimlerin elde edilmesine olanak sağlamaktadır. Geleneksel kayıp veri teknikleri içerisinde en iyi yöntem olarak görülmektedir. Seçkisiz kayıp varsayımı altında yansız parametre kestirimi yapılabilir. Bu özelliğinden dolayı stokastik regresyon ile atama yönteminin çoklu atama yöntemine ve en çok olabilirlik yöntemlerine alternatif bir yöntemdir (Enders, 2010).

Little (1992), bağımsız değişkenlerin kayıp verilerinin regresyon ile yapılan atama sonucunda elde edilen kestirimler yanlış sonuçlar doğurmaktadır. Bu yanlışlık sorunu stokastik regresyona eklenen rastgele artık terimleri tarafından çözümlenmektedir.

Baraldi ve Enders (2010), Stokastik regresyon yönteminin TSK ve SK koşullarında, regresyon ile yapılan atamaya göre yansız ve daha net kestirimler verdiğini belirtmektedir.

Beklenti - Maksimizasyon Algoritması

Beklenti-maksimizasyon ataması eldeki veriler üzerinden kurulan regresyon denklemi ile kayıp verilerin tahmin edildiği beklenti adımı ilk aşamayı ve bu adımdan sonra oluşan eksiksiz veri seti üzerinden yeniden kurulan regresyon denklemi ile ikinci aşama olan maksimizasyon adımı tekrarlı bir şekilde birbirini takip etmektedir (Enders, 2010). Acock (2005), bu yaklaşım eldeki tüm gözlenen değerler arasındaki ilişki üzerine kuruludur. Tekrarlı bir şekilde yapılan atamalar sonucunda daha fazla bilginin elde edilmektedir çünkü her atamada daha önce yapılan atamadaki bilgileri kullanmaktadır. Elde edilen sonuçlar benzer olana kadar atamalara devam edilir.

Beklenti - maksimizasyon atamasında izlenen işlem adımları yol şu şekildedir (Allison, 2003);

1. Bu yöntemde tahmin edilen ortalama, varyans ve kovaryans parametreleri liste bazında silme ve çift bazında silme yöntemleri ile elde edilen eksiksiz veriler üzerinden elde edilir.

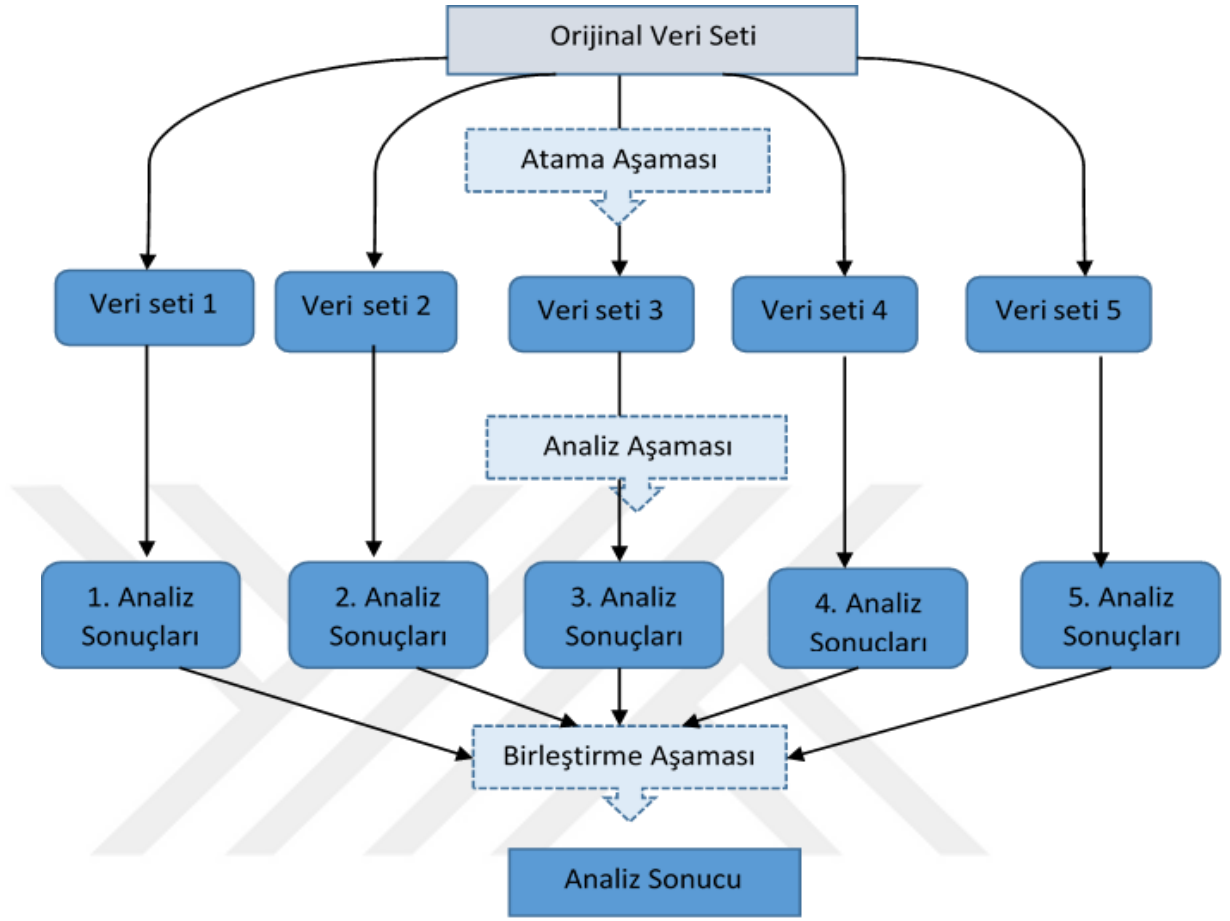
2. Kayıp verileri tahmin etmek için her kayıp veri seti için regresyon denklemi oluşturulur. Kayıp veriler tahmin edilirken ilk adımda bulunan ortalama, varyans ve kovaryans değerleri de dikkate alınır.
3. Bu oluşturulan regresyon denklemleri kayıp veri içeren tüm değişkenler için tahmini değer üretmede kullanılır.
4. Bütün değişkenlerdeki kayıp veriler için yapılan atama işleminden sonra varyans ve kovaryans değerlerindeki düzeltmeden sonra ortalama, varyans ve kovaryans değerleri yeniden hesaplanır.
5. Kestirilen değerler yakınsayana kadar yani değişmeyen kestirimler elde edilene kadar 2. Aşamaya dönülür ve 2., 3. Ve 4. Adımlar tekrarlanır.

Elde edilen parametreler tahmin etme sürecinin bir parçası için kullanılmaktadır. Bu parametrelerin sürece katılması eldeki sonuçların yanlı olmasına sebep olabilir. Parametre değerlerinin doğru üretilmesinin yanında standart hata kestirimi yapmama gibi bir olumsuz yanı vardır (Allison, 2003; Alpar, 2003).

Çoklu Değer Atama

Kayıp veri yöntemlerinde olan bu yöntem, veri setinde gözlenemeyen yani kayıp olan verilerin yerine birden fazla değer ataması yapmaktadır. Yani yapılan atamalar sonucunda araştırmacının elinde birden fazla eksiksiz veri seti olacaktır. Elde edilen bu yeni veri setleri birbirinden bağımsız bir şekilde analize alınır. Bu veri setlerinin analizleri sonucunda nihai kestirimlerin yapılabilmesi için birleştirilir.(Van Buuren,2012). Bu yöntemde önemli olan kısım yapılacak atama sayısının kaç olacağına karar vermektir. Bu atama yönteminde atama sayısının en az 2 olmakla beraber üst sınırı yoktur. Bu atama yönteminde atamanın fazla yapılması ile standart hatanın düşmesi sağlanacaktır. Bu sebeptendir ki ne kadar çok atama yapılırsa o kadar sağlıklı sonuçlar elde edilecektir (Schafer ve Olsen, 1998). Bu yöntemde analizler atama aşaması, analiz ve birleştirme aşaması olmak üzere üç aşamada gerçekleşmektedir. Atama aşamasında farklı sayıda atama yapılarak (örn: $m=5$) veri setleri oluşturulur. Oluşturulan veri setleri üzerinden ise analiz aşamasına geçilir. Analiz aşamasında her veri seti için bir tane olmakla beraber toplamda m kere uygulanmaktadır. Analizler sonucunda eldeki m adet kestirilen parametrelerin

ortalaması alınır (Enders, 2010). Atama sayısının 5 ($m=5$) olduğu durumda, çoklu değer atama yönteminin gösterimi Şekil.4'deki gibidir (Van Buuren, 2012).



Şekil 4. Çoklu Değer Atama İşlem Adımları

2.4. Klasik Test Kuramı (KTK)

Test geliştirmede kullanılan esas amaç uygun maddelerin belirlenmesidir. Bunu sağlamak için ise madde analiz çalışmalarından yararlanır (Baykul, 1979). Bir test oluşturulurken ve uygulanırken hangi test kuramının daha avantajlı olduğu üzerinde hala tartışılan bir konudur. Madde analizi çalışmalarında en çok kullanılan kuram Klasik Test Kuramıdır.

Ölçme literatürün de kişilerin performanslarının altındaki örtük özelliklerin açıklanması için geliştirilen kuramlardan birisi olan Klasik Test Kuramı (True Score Theory), psikolojik özelliklerin ölçme için kullanılmakta olan testlerden elde edilen iki bileşene dayalıdır. Bu bileşenlerden bir tanesi gerçek puan, diğeri ise hata puanıdır (Köse ve Öztemur, 2014). KTK, ölçme sorularından elde edilen puanlar üzerinden gerçek puanı tahmin etmeye çalışmaktadır. KTK, Spearman'ın basit eşitliği olan;

Gözlenen puan(X) = Gerçek puan(T) + Hata puanı(E) üzerine kurgulanmıştır (Crocker ve Algina, 1986)

$$X=T+E$$

X: Gözlenen puan

T: Gerçek puan

E: Hata puanı

Ölçme hatasına, ölçülen nesnelere gerçek değerleri ve ölçme sonuçlarından elde edilmekte olan değerlerin aralarındaki farka denilmektedir. KTK'nda hata; $E= X - T$ eşitliği ile tanımlanır (Crocker ve Algina, 1986). Tesadüfi hata olarak kabul edilen ölçmedeki tesadüfi hata, bütün ölçmeler üzerinde gerçek puandan bağımsız şekilde ve ölçmede bulunan hata puanının, başka ölçmelerdeki hata puanlarından bağımsızlığı kabul edilir. (Crocker ve Algina, 1986):

Ölçme hatası, ölçme sonuçlarının genellenebilirliğini ve ölçme sonuçlarına dayalı şekilde verilmiş olan kararların isabetliliği sınırlanır. Ölçme hatası sonuçlar üstünde hatanın türlerini ve kaynaklarını bilmek açısından da önem teşkil etmektedir (Köse ve Öztürk, 2014). Yapılan ölçme işlemlerinde deneklerde ve ölçme aracında değişme olmadığı halde, yapılan ölçümler arasında fark olması ölçmeye hata karıştığını göstermektedir. Ölçülen özelliğin ölçme sırasında gerçekte bir değeri vardır o değer özelliğin gerçek değeri(gerçek puanı), ölçme yapıldıktan sonra meydana çıkan değer ise ölçülen özelliğin gözlenen değeridir. Bu çıkan sonuç ise ölçülen özelliğin gözlenen puanıdır.

Klasik test kuramında, ölçülen denegin belirli bir değere sahip olma derecesi, o denegin o özelliği ölçmek için hazırlanmış test maddelerine verdiği cevapların ya da tepkilerin toplamı ile elde edilmektedir. Ölçme sonucunda elde edilen puan, denegin ölçülmek istenen özelliğe sahip olma derecesini göstermektedir. Eğer alınan puan yüksek ise denek ölçülmek istenen özelliğe yüksek düzeyde sahip, düşük olması halinde ise denegin ölçülmek istenen özelliğe düşük düzeyde sahip olduğunun göstergesidir. Ölçülmek istenen özellik hakkında madde düzeyinde bilgi elde etmek için madde gücü (P_j) ve ayırıcılık gücü (r_{jx}) indeksi olarak sağlanmaktadır.(Fan, 1998). Klasik test kuramında en yaygın olarak kullanılan test ve madde istatistikleri; madde ayırıcılık indeksi, madde güçlük indeksi, madde varyansı, maddeler arası ilişki

(korelasyon), madde güvenilirlik katsayısı, testin güçlüğü, ortalaması varyansı ve güçlüğüdür. Yapılan analizler sonucunda madde güçlüğüne ve ayırt ediciliğine bakılarak istenen düzeyde test güvenilirli ve geçerli bir test elde edilebilir. Bu özellikler aşağıda başlıklar halinde açıklanmaktadır.

2.4.1. Klasik Test Kuramında Madde İstatistikleri

Madde Güçlük İndeksi

Test analizinde madde ve test istatistiklerini kestirme aşamasında kullanılan başlıca iki yöntem vardır. Bunlardan birincisi klasik test teorisine, ikincisi de ve madde tepki teorisine dayanan yöntemlerdir.

Bu iki kuramında ortak varsayımı ise test puanlarının normal dağılım göstermesidir. Bu varsayım klasik test teorisine ve madde tepki teorisinin ortak yanını oluşturmaktadır. Her iki kuramda da normallik şartının sağlanması durumunda madde tepki teorisindeki a ve b parametreleri ile klasik test teorisindeki madde güçlük indeksi (p) ve madde ayırıcılık gücü indeksi (r) arasında geçiş sağlanabilmektedir. Yani, normal dağılım sağlandığında klasik teorideki madde istatistiklerinden madde tepki teorisindeki güçlük (b) ve ayırt edicilik (a) parametrelerini kestirmek mümkün olabilmektedir (Crocker ve Algina, 1986)

1 ve 0 puanları ile puanlanan bir maddenin madde puanlarının ortalamasına madde güçlük indeksi denilmektedir. Bir diğer ifade ile madde güçlük indeksi, bir tane maddeyi doğru cevaplayanların testti almakta olanların tümünün sayısına bölümü demektir.

$$P_j = \frac{D_j}{N}$$

P_j = j maddesinin güçlük indeksi

D_j = j maddesini doğru cevaplayanların sayısı

N = Toplam öğrenci sayısı (Crocker ve Algina, 1986; Baykul,2000).

Bu bakımdan, madde güçlük ölçüsü sadece maksimum performansın ölçüldüğü testler ve ölçeklerde geçerlidir. Doğru ve yanlış cevapların var olmadığı tipik davranış ölçekleri için madde güçlük ölçüsü kullanılmamaktadır (Erkuş, 2003). Yukarıda

verilen eşitlikten anlaşılacağı üzere madde güçlük indeksindeki temel işlev, problemin zorluk veya kolaylık düzeylerini barındırmaması hususudur. İndeks (p), ne kadar büyük (1'e ne kadar yaklaşır) ise madde de o kadar kolay, ne kadar küçük (0'a ne kadar yaklaşır) ise zor şeklinde yorumlanmaktadır. Ancak KTK için indeks sadece maddelerin kendi zorlukları ve kolaylıklarına göre ilişkili biçimde bir değerde değil, testleri yanıtlamakta olan kişilerin yeteneklerinin seviyelerine göre de değişmekte olan bir değerdedir. Yani bu indeks tesadüfi olarak seçilen bir kişinin o maddeyi doğru olarak cevaplama olasılığını belirtir. Maddeye cevap veren bireylerin yetenekleri yüksek ve düşük düzeyler arasında değişen bir dağılım gösterir.

Madde varyansı

Madde standart sapmasının karesine eşit olan madde varyansı, maddelerin 1 ve 0 biçimindeki puanlanmakta olduğu durumlar için o maddenin güçlük indeksi (p) ile güçlük indeksinin tersi (q) çarpımlarına eşittir (Lord ve Novick, 1968).

Madde varyansının büyümesi ile o maddeler üzerinden ölçülmesi hedeflenen özellikler açısından kişilerin aralarındaki farkların ortaya çıkma gücü artmaktadır. Varyansların, ölçülen nitelikler açısından kişiler aralarındaki farklarını belirlemekle görevli istatistikler olukları düşünüldüğünde, madde güçlük indeksi ölçmekte olduğu nitelikler açısından farkları en açık şekilde oraya koyan maddelerdir (Baykul, 2000). Ancak bu belirlemede, tamamı orta güçlük seviyesindeki maddelerden meydana gelmekte olan bir testin daima cevaplayıcıların aralarındaki farkı en iyi biçimde saptadığı anlamına gelmemektedir. Özellikle, uç yetenek düzeylerine sahip kişilerin birbirlerinden ayırt edilmeleri, kolay ve zor maddelerin de testlere dahil edilmesi ile mümkün olabilmektedir (Crocker ve Algina, 1986).

Tüm bunların yanı sıra Ebel ve Frisbie (1991), kişi puanlarının dağılımlardaki ranjı arttıracaklarını öne sürerek daha ayırt edici özellikteki ölçmeler yapılması için başarı testlerinde ağırlıklı şekilde güçlük düzeylerinde yer alan maddelere yer verilmesi gerekliliğini önermektedirler.

Madde ayırıcılık indeksi

Bir testteki veya ölçeklerdeki maddeler, madde ile ölçülmesi hedeflenen özelliklere sahip olanlar ve olmayanlar için birbirinden ayırt edilmelidir. Bu özellikleri ile maddeler, maddenin ayırt edici gücü ile ifade edilmektedir. Maddenin ayırt edicilik

gücü, yine maddenin geçerliliği olarak da bilinmektedir. Maddelerdeki ayırt edicilik gücünü belirlemek üzere belirleyiciler için çeşitli yöntemler mevcuttur. Bu amaca yönelik olarak ulaşılan indekse madde ayırmacılık indeksi (r_{jx}) denilir. Maddenin ayırt ediciliğinin saptanabilmesi için bir ölçüt gereklidir. Bu amaca yönelik olarak belirlenen ölçüt, o maddenin bulunduğu test veya ölçeğe dair istatistiklere dâhil ise buna “iç ölçüt” denilir. Ölçülmesi hedeflenen özelliği ölçebilen, geçerlilik ve güvenilirliği kanıtlanmış bir başka ölçme aracı ile ulaşılan istatistiklere ise “dış ölçüt” denilmektedir (Baykul, 2000; Erkuş, 2003).

İç ölçüt kullanımının daha yaygın tercih edildiği madde ayırmacılık indeksi hesaplama yöntemlerinde ölçeklerden elde edilen toplam puanlar dikkate alınmaktadır. Toplam puanların maddenin geçerliliğinde ölçüt alınmasının temel nedeni, denenmekte olan maddelerin, testle ölçülmesi istenen özelliğini geçerli olarak ölçebildiği sayılıdır (Turgut ve Baykul, 2012). İç ölçüt kullanılarak madde ayırt edicilik gücünün hesaplanmasında, “madde toplam test korelasyon katsayıları”, “alt-üst %27’lik grup ortalamalarının bağımsız gruplar t-testiyle karşılaştırılması”, “regresyon analizi” gibi yöntemlerin kullanımı tercih edilmektedir (Erkuş, 2003).

Madde ayırmacılık indeksi -1 ve +1 aralıklarının da değişmektedir. Negatif değerlerin maddelerin düşük yetenekli kişiler tarafınca büyük oranlarda doğru yanıtlanırken yetenekli kişiler tarafınca doğru cevaplanmaması durumları ile ortaya çıkmaktadır. Ayırmacılık indeksinin negatif olduğu bir maddede ters işleyen sorunlu bir madde özelliği mevcuttur. İndexte pozitiflik ve 0,25’in üstünde bir değer beklenmektedir. Ayırt ediciliğin düşük olduğu maddelerin tekrardan yazılması veya testten çıkarılması gereklidir (Courville, 2004).

Ayırmacılık indeksleri örneklem büyüklüklerinden ve örneklemde bulunan kişilerin yetenek ranjından etkilenmekte olan bir ölçüdür. Yetenek seviyeleri açısından homojenlik gösteren ve küçük bir örneklem üstünden hesaplanabilen madde ayırmacılık indeksleri yanlış sonuçlar verebilmektedir (Ebel ve Frisbie, 1991).

Bu açıdan, küçük gruplarda yapılmaktaki ölçmeler ile yaygın şekilde kullanılmakta olan KTK ‘ya dayalı analizlerde örnekleme hastasından kaynaklanan problemlerle karşılaşmadığı söylenebilir. Küçük ölçekli bir örneklem üzerinden ulaşılan veriler ile yapılan analiz üzerine ayırt ediciliği yüksek saptanan bir madde farklı küçük bir örneklem ile analiz edildiği zaman ayırt ediciliğin düşüklüğü hatta

negatifliđi ortaya çıkmaktadır (Ebel ve Frisbie, 1991). Bu indeksin belirlenmesinde korelasyon kullanılır. (Baykul, 2000)'a göre korelasyon iki deđişkenin birlikte deđişimin ölçüsüdür. Madde ayırıcılık gücü indeksi, test puanları ile madde puanları arasındaki korelasyon olarak tanımlanır. Madde ayırıcılık gücü indeksi hesaplanmasında çift serili korelasyon katsayısı (biserial), nokta çift serili korelasyon katsayısı (point biserial) ve tetrakorik korelasyon katsayısından yararlanır.

Çift serili korelasyon katsayısı deđişkenlerden biri normal dağılım gösteren ve sürekli, diđeri ise gerçekte sürekli iken yapay olarak süreksiz hale getirilmiş (iki kategorili) madde ve test puanları arasındaki ilişkinin hesaplanmasında yararlanır. (Ferguson, 1976; Howell, 1992). Test puanlarının sürekli, madde puanlarının da normal dağılım ve sürekli iken, madde puanları 1 ve 0 olarak puanlama yapılarak süreksiz hale getirilerek test puanları ile arasındaki ilişki çift serili korelasyon ile hesaplanır. "r_ç" ya da "r_b" ile gösterilir.

$$r_b = \frac{\bar{x}_{jD} - \bar{x}_x}{s_x} \cdot \frac{p \cdot q}{y}$$

r_b = Madde ayırıcılık gücü indeksi

\bar{x}_{jD} = j maddesini doğru cevaplayanların ortalaması

\bar{x}_x = Test puanlarının ortalaması

s_x = Testin standart sapması

p_j = j maddesinin doğru cevaplandırılma yüzdesi

q_j = j maddesinin yanlış cevaplandırılma yüzdesi

y = p ve q oranları ayırım noktasında normal dağılım üzerindeki ordinat yüksekliđi

Pearson momentler çarpımı korelasyonunun özel bir hali olan nokta çift serili korelasyon kat sayısı -1 ve +1 arasında değerler alır (Arıcı,1991; Baykul,2000) Yapılan araştırma değişkenlerden biri normal dağılım gösteren sürekli değişken, diğer değişkenin ise gerçek süreksiz (iki kategorili) olduğu durumda değişkenler arasındaki ilişkiyi bulmak için nokta çift serili korelasyon katsayısı kullanılır (Ferguson, 1976; Howell, 1992). Örnek vermek gerekirse nokta çift serili korelasyon katsayısı medeni durum (evli- evli değil) ve gelir düzeyi; cinsiyet (kadın-erkek) ve akademik başarı; madde puanlarının gerçekte 0 ve 1 gibi gerçek süreksiz bir değişken olduğu ve test puanlarının sürekli ve normal bir dağılım gösterdiği sayılıta göz önüne alındığı durumlarda kullanılır. “ r_{nc} ” ya da “ r_{pb} ” ile gösterilir.

$$r_{pb} = \frac{\bar{x}_{jD} - \bar{x}_x}{s_x} \cdot \sqrt{\frac{p_j}{q_j}}$$

r_{pb} = Madde ayıricılık gücü indeksi

\bar{x}_{jD} = j maddesini doğru cevaplayanların ortalaması

\bar{x}_x = Test puanlarının ortalaması

s_x = Testin standart sapması

p_j = j maddesinin doğru cevaplandırılma yüzdesi

q_j = j maddesinin yanlış cevaplandırılma yüzdesi

Tetrakorik korelasyon katsayısı, değişkenlerin sürekli ve normal dağılım gösterdiği halde yapay olarak iki kategorili (dikatom) hale getirilmesi durumunda, yani dikatom değişkenlerin örtük sürekli değişkenleri yansıttığı durumlarda kullanılır. Örneğin; deneklerin verdiği cevapların (0:yanlış, 1:doğru) sürekli bir değişkenin yansıması olarak düşünülmesi gereken durumlarda tetrakorik korelasyon kullanılır (Uebersax, 2015; Baykul, 2000). Tetrakorik korelasyon katsayısının hesaplanmasında aşağıdaki eşitlikten yararlanılmaktadır. r_t veya p_t ile gösterilir.

$$r_t = \text{Cos} \left(\frac{180^\circ}{1 + \sqrt{\frac{B.C}{A.D}}} \right)$$

Verilen eşitlikte r_t korelasyon katsayısını gösterir. Eşitlikteki $\frac{B.C}{A.D}$ oranı hesaplandıktan sonra çıkan katsayı, kosinüs fonksiyonu bulunan hesap makinası veya trigonometri cetvelinden yararlanılarak bulunabilir (Baykul, 2000).

Madde güvenilirlik katsayısı

Güvenirlik bir ölçme aracının doğruluğu veya duyarlılığı anlamına gelir. Güvenirlik indeksleri, belli bir ölçme aracından elde edilen puanların ne derece tutarlı ve tekrarlanabilir olduğu hakkında bir fikir verir. Bir testin neyi ölçtüğünü değil, ölçtüğü şeyi ne kadar doğru ölçtüğünü sorarız. Elde edilen puanın doğruluğu nedir? Bir bireyi tekrar ölçtüğümüzde, yeniden elde edilen puan ne derece tutarlı olacaktır? (Thorndike, 1982). Klasik test teorisinde güvenilirliğin hesaplanmasında farklı yöntemler vardır. Bu yöntemler test tekrar test, paralel testler, yarılama, Cronbach's Alpha(α), KR-20 ve KR-21 olarak sıralanabilir.

KR-20 ve KR-21 formülleri testteki her bir maddenin aynı değişkeni ölçtüğü, yani testin ölçtüğü şeyin homojen olduğu sayılına dayanır. KR-20 formülü genellikle iki maddeli (binary) testlerde kullanılır. Yani doğruların 1 yanlışların 0 ve boş bırakılan maddelere ise hiç puan verilmeyen testlerde kullanılır. Eğer testteki maddeler farklı ağırlıklarla puanlanmışlarsa KR-20 kullanılmaz. KR-20 testteki maddelerin iç tutarlık katsayısını verir (Tekin, 2017).

Baykul (2000)'a güvenilirlik katsayısı, paralel olarak ölçülen iki ölçme arasındaki katsayıdır. Eğer bir test k tane alt testten oluşuyor ise Cronbach's α , bu araştırmada olduğu gibi tek bir testten oluşuyor ise Cronbach's α nın özel hali olan KR-20 güvenilirlik katsayısından yararlanır. KR-20 testi oluşturduğu maddeler arasındaki iç tutarlılığı vermektedir. KR-20 0,00-1,00 arasında değerler alır. KR-20'nin 1,00 a yaklaşması iç tutarlılığın yüksek aynı özelliği ölçtüğünü, 0,00'a yaklaşması ise iç tutarlılığın düşük farklı özellikleri ölçtüğünün göstergesidir. KR-20 güvenilirlik katsayısının aşağıdaki eşitlik yardımıyla bulabiliriz.

$$KR-20 = \left(\frac{K}{K-1} \right) \left[1 - \frac{\sum_{j=1}^k P_j \cdot q_j}{s^2 t} \right]$$

K= Testteki madde sayısı

Σ =Toplam

P_j = Maddeyi doğru cevaplayanların oranı (j. maddenin güçlük indeksi)

q_j =Maddeyi doğru cevaplamayanların oranı (1- P_j)

s²=Test puanlarının varyansı

2.4.2. KTK'nın Avantajları ve Dezavantajları

KTK'nın sağlamış olduğu avantajlar aşağıdaki gibi sıralanabilir (Hambleton ve jones, 1993; Fan, 1998; Stage, 2003):

- Okullardaki başarı testlerini geliştirmiş olan öğretmenler ve yapmış oldukları araştırmalar bakımından çeşitli ölçeklerin kullanımına yönelik bilgi toplanılan ancak ölçme alanlarında üst düzeyde bilgiye sahip olmayan araştırmacılara yönelik olarak KTK, madde ve test istatistikleri hesaplanması açısından daha çok tercih edilmektedir.
- Karmaşık analiz süreçlerini basite indirgeyen bilgisayar yazılımlarının çeşitli yeni yaklaşımların yaygınlaşmasına imkân vermektedir. Bu açıdan KTK için hesaplama kolaylıkları ve yorumlama kolaylıklarına yönelik avantajların zamanla avantajdan çıkabilir.
- KTK'da örneklem büyüklüğü bakımından katı bir duruş bulunmamaktadır. Madde ayırıcılık indeksi gibi istatistikler daha doğru yöntemler ile hesaplanabilmek açısından küçük ve homojen örneklemeler uygun olmasa da kuramın matematiksel alt yapısındaki basitlikten dolayı, çok büyük örneklemelere gereksinim duyulmamaktadır.
- Özellikle MTK'nın kullanımına imkân sağlamayan veri yapıları için KTK başarılı sonuçlar elde etmektedir. Literatür incelendiğinde bu her iki kuramın büyük örneklemeler için karşı karşıya kaldıkları durumlarda hem madde hem birey özellikleri açısından benzer ve tutarlı sonuçlar

verdiklerine kanıt sunabilen arařtırmalar yer almaktadır. KTK'nın saęlamıř olduęu dezavantajlar ařaęıdaki gibi sıralanabilir ; (McDonald ve Paunonen 2002; Fan, 1998; Nartgün, 2002; Embretson ve Reise, 2000).

- KTK'ya baęlı olarak geliřtirilmekte olan testlerde ve ölçeklerdeki madde istatistikleri, testlerin geliřtirildięi gruplardan geliřtirilmekte ve grup deęiřtięi zaman ulařılan istatistiki deęerler de deęiřmektedir.
- Madde örneklemine baęlı olan test puanları ve birey örneklemine baęlı madde istatistikleri mevcuttur. Eęer testler göreceli biçimde kolay maddelerden meydana geliyorsa, kiřilerin testlerden aldıkları puanlar yüksek; testler göreceli biçimde zor maddelerden meydana geliyorsa kiřilerin testlerden aldıkları puanlar düşük görülecektir. Bu bakımdan, KTK'için kiřilerin başarı seviyeleri de testlerde bulunan maddelere baęlı biçimde görecelidir. Aynı zamanda kiřiler ve test istatistikleri, birbirlerine baęlı oldukları için Bilgisayar Ortamında Bireyselleřtirilmiř Test (BOB test) uygulamalarında teorik zorluklar ortaya çıkmaktadır.
- KTK modellemesi üzerinde bütün hatalar tesadüfi hata kabul edilmektedir. Bu duruma karřılık kültürel alt yapı ve cinsiyet gibi deęiřkenlerden kaynaklanan sistematik hataların ölçmeye karřılabilmektedir. Testlerin yanlılıęına yol açan bu gibi hatalar, KTK'da genel olarak görüldüğü edilebilir.
- KTK için ölçümlerin güvenilirlikleri testleri yanıtlamakta olan kiřilerin özellikleri, testlerin özellikleri ve test uygulayıcıların kararlarına ve yöntemlerine göre deęiřmektedir.
- KTK için ölçme hataları ile güvenilirlik katsayısı bütün gruplar için hesaplanmaktadır. Çeřitli yetenek seviyelerindeki kiřiler açısından ölçme araçları aynı seviyelerde güvenilirliklere sahip olmayacaęından bu durum bir dezavantajı temsil etmektedir.
- KTK için geliřtirilen uygulamalarda testlere katılmakta olan kiřilere verilebilen en önemli bilgi, onların bütün testler üzerinde gösterdikleri performanslarını yansıtabilen toplam test puanlarıdır. Kiřilerin çeřitli

maddeler üstünde sergiledikleri performanslar KTK'da detaylı şekilde ele alınmamaktadır.

2.5. İlgili Araştırmalar

Enders (2003), farklı kayıp veri yöntemlerinden liste bazında silme, çift bazında silme, ortalama değer atama, bireysel ortalama atama ve beklenti-maksimizasyonu yöntemlerini Cronbach-alfa güvenilirlik katsayı kestirimleri bağlamında karşılaştırmıştır. Yapılan çalışmada örneklem büyüklükleri ($n=100$, $n=300$ ve $n=500$), madde sayıları ($k=10$ ve $k=20$), ölçekteki madde cevapları ($c=3$, $c=5$ ve $c=7$) ve kayıp veri oranları (%15 ve %30) olacak şekilde manipüle edilerek Cronbach-alfa güvenilirlik kestirimlerindeki yanlılık, güven aralığı ve ortalama hataların karekökleri temelinde incelemiştir. Analiz sonucunda çift bazında silme ve beklenti maksimizasyonu yöntemlerinin diğer yöntemlere göre daha yüksek değerler üretmişlerdir. Kayıp veri oranlarına göre ise tüm yöntemlerin negatif yönlü kestirimler üretmiştir. TSK bağlamında beklenti-maksimizasyon yönteminin örneklem büyüklüğü ve madde sayısının artmasıyla paralellik gösterip daha iyi sonuçlar ürettiği fakat liste bazında silmenin ise negatif yanlı sonuçlar verdiği görülmüştür.

Enders, (2004), 1000 kişilik yapay veri seti üzerinden SK ve SOK koşulları altında farklı kayıp veri baş etme yöntemlerinin Cronbach-alfa kestirimleri üzerindeki etkisi incelemiştir. Kullanılan kayıp veri yöntemlerinden TSK koşulu altında liste bazında silme, çift bazında silme, ortalama atama ve beklenti-maksimizasyon yöntemleri ile negatif yanlı güvenilirlik kestirimleri; SK koşulu altında ise beklenti-maksimizasyonun yansız fakat diğer yöntemlerin ise negatif yönlü kestirimler üretmiştir. SK koşulu altında, evrene ilişkin güvenilirlik değerine en yakın sonuçların beklenti-maksimizasyon yöntemiyle elde edilmiştir. Güven aralığına yönelik kestirimlerde her iki kayıp veri koşullarında da beklenti-maksimizasyon atama yönteminin daha etkili olduğu görülmüştür. SK koşulu altında liste bazında silme ve çift bazında silmenin kabul edilebilir değerler altında kestirimler ürettiği fakat TSK koşulu altında isabetli kestirimler üretmiştir. Ortalama atama yöntemi ise her iki koşulda isabetsiz kestirimler üretmiştir.

Acock (2005), farklı kayıp veri yöntemi üzerinden bir çalışma yapmıştır. Bu çalışmada gerçek veriler üzerinden analizler yapılmıştır. Yapılan bu çalışmanın veri seti 2002 yılında sosyal araştırma için 818 kişiye uygulanan bir ölçeğe ait verilerdir.

Yapılan analizlerde eğitim düzeyi, yaş, gelir düzeyi, mutluluk düzeyi ve çocuk sayısı dikkate alınarak yapılmıştır. Yapılan analizlerde TSK varsayımı dikkate alınarak analizler yapılmıştır. Veri setinden %50 kayıp veri içeren veri seti ile %40 seçkisiz kayıp veri içeren iki veri seti oluşturmuştur. Eksiksiz veri seti ve oluşturduğu iki veri setleri üzerinden analizler yapılmıştır. Kayıp veri yöntemlerinden liste bazında silme, çiftler bazında silme, ortalama atama, beklenti maksimizasyonu ve çoklu değer atamaları yapmıştır. Elde edilen regresyon kestirimlerini ve hata oranlarını incelemiştir. TSK varsayımı altında yapılan analizler sonucunda ortalama atamanın en kötü sonuçları verdiği ulaşılmıştır.

Chen, Wang ve Chen (2011), tarafından yapılan bir araştırmada TSK koşulu altında farklı oranlarda kayıp veri içeren veri setleri incelenmiştir. Yapılan bu araştırmada açıklayıcı ve doğrulayıcı faktör analizi kestirimlerini etkilemede; kayıp veri atama yöntemlerinden liste bazında silme, regresyonla atama, beklenti-maksimizasyon, seri ortalama atama, doğrusal eğilim ve doğrusal interpolasyon çerçevesinde değerlendirmişlerdir. Yapılan açıklayıcı faktör analizinde açıklanan varyans oranları %54 - %99,16 arasında değiştiği, kullanılan kayıp veri yöntemlerinin birbirine yakın sonuçlar ürettiği fakat beklenti-maksimizasyon yönteminin diğerlerine göre daha yüksek performans göstermiştir.

Çokluk ve Kayri (2011) , kayıp verilerin olmadığı veri setleri ile kayıp verilere yaklaşık değer atama yöntemlerinin kullanıldığı veri setlerinin geçerlilik ve güvenilirlik bağlamında incelemiştir. Yapılan çalışmada kayıp veri ile baş etme yöntemlerinin, veri setinin faktör yapısına, madde-toplam korelasyonuna ve Cronbach güvenilirlik katsayısına etkisi incelenmiştir. 200 kişiden oluşan veri setinin seçkisiz yöntemle %15-%20 ve %0-%50 arasında eksiltmelerle kayıp veri içeren yeni veri setleri elde edilmiştir. Bu veri setlerinin de orijinal veri seti gibi tek faktörlü yapı oluşturduğu gözlenmiştir. Bu atamalardan yaklaşık değer atama yönteminin orijinal veri setine daha yakın sonuçlar verdiği görülmüştür. Bu üretilen veri setlerinin orijinal veri setine göre açıklanan varyans oranının da, faktör yüklerinde ve güvenilirlik katsayısının da düşüşe sebep olduğu görülmüştür. Tek faktörlü yapının bütün veri setlerinde korunduğu görülmüştür.

Hohensinn ve Kubinger (2011), İki kategorili(1-0) puanlanan simülatif veriler üzerinden madde uyumu ve model geçerliliği arasındaki ilişkiyi incelemiştir. Yapılan

analizler üç ayrı veri seti üzerinde yapılmıştır. Birinci veri setinde kayıp verilerin yerine hepsi yanlış olarak değerlendirilerek veri setinde 0 olarak kodlanmıştır. İkinci veri seti kayıp veri içermemektedir. Üçüncü veri seti ise seçkisiz bir şekilde kayıp veriler oluşturularak analiz dışında bırakılmıştır. Bu durumda üç farklı veri seti üzerinden analizler yapılmıştır. Ve yapılan analiz sonucunda madde ve yetenek parametrelerinde manidar farklılıklar elde etmiştir. Testin son maddelerine ulaşamaması durumunda TSK ve SK varsayımlarının ihlal edildiğini ispatlamışlardır.

Demir (2013), kayıp veri içeren iki kategorili puanlanan SBS 2011 Matematik testindeki veri setinin kayıp veri yöntemlerine göre psikometrik özelliklerini incelemiştir. Bu çalışmada kayıp veri ile baş etme yöntemlerinden 12 yöntemle başvurulmuştur. Bu yöntemler LBS, BA, AÇO ve ÇA yaklaşımına dayalı yöntemlerdir. Yapılan analizler sonucunda elde edilen kestirimlerde dizin silme ve sıfır atamanın iki kategorili puanlanan veri setleri için uygun olmadığını belirtmiştir. Basit atama yönteminin yanlış sonuç verdiği ve en çok olabilirlik ve çoklu atama yöntemlerinin iki kategorili puanlanan veri setleri için daha uygun olduğunu belirtmiştir. Seçkisiz olmayan kayıp verilerin varlığında ikili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin kestiriminde uygun kayıp veri yönteminin gerekli olduğunu belirtmiştir.

Akbaş ve Tavşancıl (2015), kayıp veriler ile baş etme yöntemlerinin ölçeklerin psikometrik özelliklerinin belirlenmesinde bağlamında incelemiştir. Bu çalışma üretilen yapay veri setleri üzerinden yapılmıştır. Bu çalışmada örneklem büyüklüğü, kayıp veri oranı ve madde sayısı farklılaştırılarak çalışma genişletilmiştir. Yapılan çalışma liste bazında silme, Öklid uzaklığı üzerinden benzer tepki örüntüsüne dayalı atama, stokastik regresyonla atama, beklenti-maksimizasyon atama ve çoklu değer atama yöntemleri kullanılmıştır. Yapılan kestirimler göz önüne alındığında liste bazında silme yönteminin tamamen seçkisiz kayıp ve kayıpların az olduğu veri setleri dışında kullanılmaması gerektiğini belirtmiştir. Çoklu değer atamanın diğer yöntemlere göre daha üstün bir yanı olmakla birlikte, veri setlerinde kullanılacak en etkili sonucu veren yöntemin belirlenemeyeceği görüşüne varmıştır.

Kürşad ve Nartgün (2015), PISA 2012 Türkiye uygulamasına katılan öğrencilerin verileri üzerinden çalışma yapmıştır. Kayıp veri ile baş etme yöntemlerinin geçerlilik, güvenilirlik ve betimsel istatistiklerinden elde edilen

kestirimler üzerinden karşılaştırma yapmıştır. Farklı büyüklüklerdeki (%5, %10 ve %20) kayıp veri oranları setleri üzerinden liste bazından silme, yaklaşık değer ataması altında; seri ortalaması, yakın noktaların ortalaması, yakın noktaların medyanı, doğrusal değer kestirimi, beklenti-maksimizasyonu, regresyon ataması ve çoklu değer atama yöntemlerini kullanmıştır. Orijinal ve kayıp veri ataması yapılan veri setleri üzerinden geçerlik ve güvenilirlik kestirimleri yapmıştır. Liste bazında silme yönteminin referans değerlere uzak bir sonuç verdiği gözlenmiştir. Tüm kayıp veri yöntemlerinde referans değerlere en yakın sonucu beklenti-maksimizasyonu, regresyon ataması ve çoklu değer atama yöntemleri vermiştir.

Kayıp veri sorununa yönelik yapılan çalışmalar incelendiğinde genellikle araştırmaların çoğu yapay veriler üzerinden gerçekleştirilmiştir. Gerçek veriler ile ilgili yapılan bir çalışma olması ile literatüre katkı sağlayacaktır.

İki kategorili puanlanan maddelerde oluşan testler de kayıp veri ile başa çıkma çok kategorili puanlanan maddelerden oluşan testlere göre, madde karakteristiğinin doğal bir özelliği olarak daha zordur. Bu sebeple çok kategorili testler için daha fazla yöntem geliştirildiği görülmektedir. Ayrıntılı analizler öncesinde kayıp veri mekanizması ve örüntüsü için ayrıntılı inceleme gerekmektedir. Elde edilen puanların anlamlılığı için tek boyutluluk dikkate alınmalıdır.

ÜÇÜNCÜ BÖLÜM

YÖNTEM

Bu bölümde, araştırmanın türü, çalışma grubu, verilerin toplanması ve verilerin analizi başlıkları yer almaktadır.

3.1. Araştırma Modeli

Bu araştırmada, ABİDE 2016 matematik alt testinin, çoktan seçmeli maddelerinin KTK' ya dayalı yöntemler ile analiz edilmesi sonucu elde edilen madde parametreleri arasındaki benzerlik ve farklılıkların tespit edilmesi amaçlanmaktadır. Bu araştırma evrene genelleme yapılmadığı ve durum belirlemeye yönelik olduğundan betimsel türde "temel araştırma" niteliğinde olduğu söylenebilir.

3.2. Evren ve Örneklem

Yapılan bu araştırma 2016 yılında Türkiye'nin 81 ilinde uygulanan ABİDE programında yer alan Matematik alt testinin çoktan seçmeli maddeleri yer almaktadır. ABİDE örnekleminin belirlenmesinde MEB Strateji Geliştirme Başkanlığından alınan okul ve şube bilgileri kullanılmıştır. 16.488 ortaokul ve 48.679 şubeyle başlanan çalışmaya, özel eğitime ihtiyaç duyan öğrencilerin bulunduğu okul ve şubeler, çalışmanın dışında tutularak 16.118 okul ve 48.091 şube üzerinden devam edilmiştir. 81 il için yapılan çalışmada her bir il için yaklaşık 400 öğrencilik gözlemin yeterli olacağına karar verilmiştir (ABİDE,2016). Bu çalışmada MEB Strateji Geliştirme Başkanlığından alınan iki kategorili olarak (1-0 şeklinde) 1000 kişilik örneklem veri seti üzerinden yürütülmüştür.

3.3. Veri Toplama Araçları

Bu çalışmada veri toplama aracı olarak ABİDE 2016 sınavında kullanılan ilköğretim sekizinci sınıf öğrencilerine uygulanan matematik alt testi için hazırlanan düzey belirleme testleri üzerinden yürütülmüştür. ABİDE 2016 uygulamasının kitapçık türlerine göre madde sayıları tabloda Tablo 4'de verilmiştir.

Tablo 4. Test Formları ve Madde Sayıları

A Formu	B Formu	C Formu
9 + 9 = 18 madde	9 + 9 =18 madde	9 + 9 =18 madde
A1: 18+2 pilot=20 madde	B1: 18+2 pilot=20 madde	C1: 18+2 pilot=20 madde
A2: 18+2 pilot=20 madde	B2: 18+2 pilot=20 madde	C2: 18+2 pilot=20 madde
A3: 18+2 pilot=20 madde	B3: 18+2 pilot=20 madde	C3: 18+2 pilot=20 madde
A4: 18+2 pilot=20 madde	B4: 18+2 pilot=20 madde	C4: 18+2 pilot=20 madde

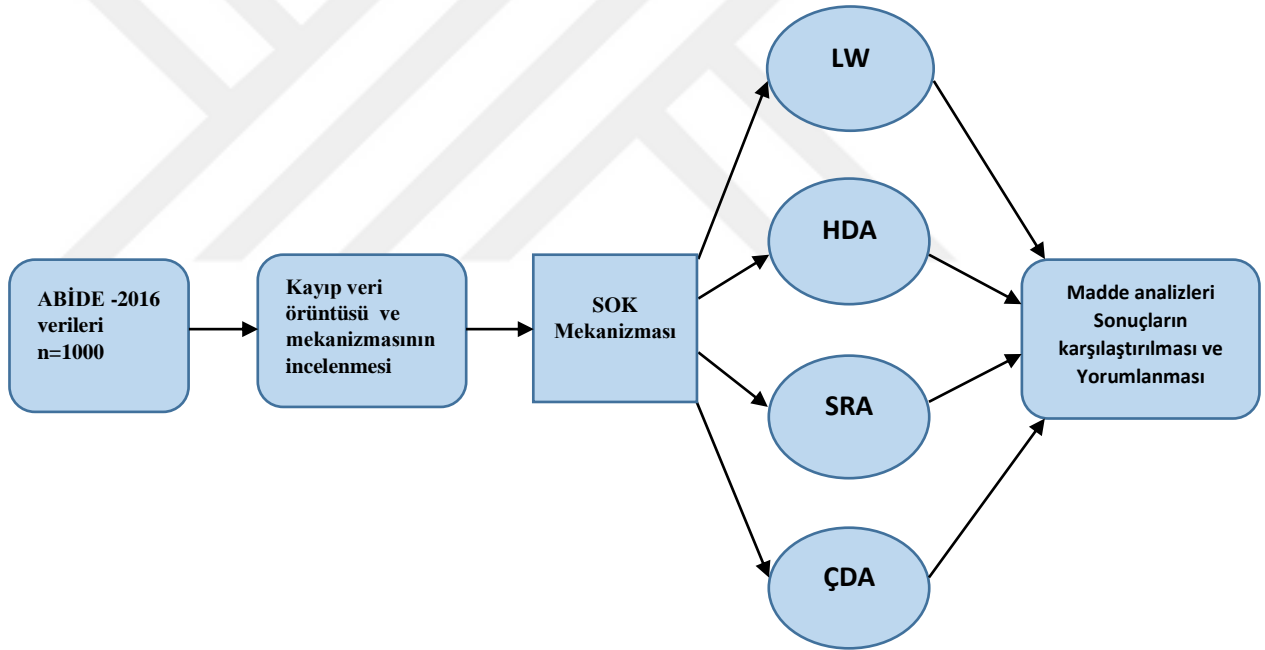
Kaynak: ABİDE 2016 Raporu

Yapılan bu çalışmada ABİDE 2016 programının matematik alt testinin A kitapçığı incelenmiştir. A1'den A4'e kadar bütün A kitapçıklarındaki maddelerin ilk 18 sorusu ortaktır. Yalnızca, son iki madde ilk kez bu uygulamada kullanılmıştır. Geriye kalan 18 maddenin dokuzu çoktan seçmeli diğer dokuzu ise açık uçlu maddelerdir. A kitapçığındaki dokuz çoktan seçmeli madde araştırma kapsamına

alınmıştır. Araştırmada kullanılan çoktan seçmeli maddeler de doğru olanlar (1) ve yanlış olanlar (0) olarak puanlanmıştır. Abide 2016 Matematik Testi A Kitapçığına yönelik veri seti Milli Eğitim Bakanlığı Ölçme ve Değerlendirme Daire Başkanlığı ile yapılan resmi yazışmalar sonucunda elde edilmiştir. Söz konusu veri seti 1000 katılımcı, 9 değişken ve A-B-C-D şeklinde kodlanan yanıtlar ile cinsiyet değişkenini içermektedir. Öncelikle ham veride A-B-C-D şeklinde kodlanan cevap örüntüleri 1-0 örüntüsüne dönüştürülmüştür. Bu dönüşümde doğru yanıtlar 1, yanlış yanıtlar ise 0 olarak puanlanmıştır.

3.4. işlem

Yapılan bu araştırmada, 1000 kişiden oluşan ABİDE-2016 Matematik testi A kitapçığının çoktan seçmeli 9 maddesinin incelenmesi için işlem aşamaları Şekil 5'teki gibidir.



Şekil 5. Araştırmada İşlenen İşlem Yolu

1. İlk aşamada veri seti incelenmiş bu veri setinin dağılımlarına yönelik betimsel istatistikler, kayıp veri örüntüsü ve mekanizması incelenmiştir.

2. Yapılan incelemeler sonucunda veri örüntüsünün SOK mekanizması olduğu kanaat getirilmiştir. Ve çalışma SOK mekanizması bağlamında incelenmiştir.

3. SOK mekanizması üzerinden kayıp veri atama yöntemlerinden LW, HDA, SRA ve ÇDA ile kayıp veriler atanmıştır. Ve bu kayıp verilerin atanması ile oluşan veri

setlerinin betimsel istatistikleri, madde analizleri ve açılımlayıcı faktör analizleri kestirimleri yapılmıştır.

4. Son aşamada bu atama yöntemlerinden elde edilen betimsel istatistikler, madde analizleri ve açılımlayıcı faktör analizi sonuçları karşılaştırılmıştır.

3.5. Verilerin Analizi

Abide 2016 Matematik Testi A Kitapçığı verilerine yönelik analizlerde ilk olarak, kayıp veri örüntüsü incelenmiştir. Verilere ilişkin incelemede, doğru ve yanlış yanıtları ile kayıp verileri içeren madde sayılarının dağılımları dikkate alınmıştır. Yapılan dağılım sonucunda dağılıma yönelik betimsel istatistikler incelenmiştir. Ayrıca en fazla ve en az görülen yanıt örüntüleri incelenmiştir. Yapılan bu inceleme sonucunda kayıp veri mekanizmasına yönelik analizler yapılmıştır. Kayıp veri mekanizmasının; Tümüyle Seçkisiz Kayıp (TSK/Missing Completely At Random-MCAR), Seçkisiz Kayıp(SK/Missing At Random-MAR) ve Seçkisiz Olmayan Kayıp (SOK/Missing Not At Random-MNAR) varsayımlardan hangisini karşıladığının belirlenmesi amaçlanmaktadır. Bu çalışmada kayıp veri mekanizması ve cevap örüntüleri dikkate alınarak analizler arasında manidar bir fark olup olmadığı dikkate alınmıştır. Yanıtlayıcıların, doğru ve yanlış cevapları ile kayıp veri içeren madde sayıları arasında manidar düzeyde bir farklılık olup olmadığı incelenmiştir. Tüm yanıtlayıcıların doğru ve yanlış cevapları ile kayıp veri içeren maddelerin sayılarının dağılımları incelenmiş ve değerlendirilmiştir. Yapılan incelemeler sonucunda elde edilen veriler, tüm yanıtlayıcıların yanıt örüntüleri ile madde ve cinsiyet gibi kategorik değişkenler düzeyinde ayrı ayrı değerlendirilmiştir.

Bu çalışmada kullanılacak kayıp veri yöntemleri belirlenmiş, iki kategorili maddelerin puanlanmasında kullanılan farklı kayıp veri yöntemi vardır. Bu yöntemlerden en belirgin olanları Liste Bazında Silme (LBS), Hot Deck Atama, Stokastik Regresyonla Atama (SRA) ve Çoklu Değer Atama (ÇDA) dır.

İki kategorili puanlanan farklı kayıp veri yöntemlerine göre yapı geçerliliği ve tek boyutluluk için açılımlayıcı faktör analizi (AFA) yapılmış ve analiz sonuçları değerlendirilmiştir. AFA kullanılarak maddelerin tek bir boyut altında toplanıp toplanmadığının belirlenmesi amaçlanmıştır.

Açımlayıcı faktör analizi sonucunda yapılan çıktılar sonucunda Kaiser-Mayer-Olkin Örneklem Büyüklüğü Uygunluk Testi (KMO) ve Bartlett Küresellik Testi incelenmiştir. Kullanılan kayıp veri yöntemlerinin analiz sonuçlarında bütün KMO değerleri .79 üzerinde elde edilmiştir. Bu sonuç atama yapılan veri setlerinin faktör analizi için uygunluğunun kanıtıdır. Bartlett testine yönelik olarak yapılan kestirimler $p < 0.05$ olarak tüm veri setlerinde elde edilmiştir.

Daha sonra LW, HDA, SRA ve ÇDA yöntemlerine göre atama yapılan veri setlerinin test ve madde parametreleri Klasik Test Kuramı temelinde ele alınarak bulunmuştur. KTK temelinde iki kategorili puanlanan veri setleri için test ve madde parametre kestirimleri yapılmıştır. Bu kestirimler madde güçlük indeksi, madde ayırt edicilik indeksi ve güvenilirlik katsayısı, minimum maksimum puanlar, test varyansı, standart sapma ve basıklık, çarpıklık katsayıları incelenmiştir.

Bütün bu analizlerin yapılmasında farklı istatistiksel analiz programları kullanılmıştır. Bütün analizleri yapmak için uygun bir istatistik yazılım programı bulunmamaktadır. Bu nedenle verilerin analizlerine uygun olarak SPSS 23.0, Factor 11.0 ve R 3.6.3 paket programlama yazılımları kullanılmıştır.

SPSS 23.0 yazılımında kayıp veriler ile başa çıkmada silmeye dayalı ve basit atama yöntemleri dışında çoklu değer atama yöntemleri bu yazılım üzerinden gerçekleştirilmiştir. Hot-deck atama yönteminde veri setinin elde edilmesi için R 3.6.3 paker programı kullanılmıştır. Kayıp veri yöntemlerine göre iki kategorili puanlanan veri setlerinin yapı geçerliliği ve tek boyutluluğu için yapılan açımlayıcı faktör analizi için FACTOR 11.0 programı kullanılmıştır.

DÖRDÜNCÜ BÖLÜM

BULGULAR ve YORUMLAR

Bu bölümünde araştırmanın amaçları doğrultusunda elde edilen analizlere, çıktılara ve elde edilen bulgulara sırasıyla yer verilmiştir.

4.1. Kayıp veri Örüntüsü ve Mekanizması

2016 ABİDE Matematik Testi A kitapçığının çoktan seçmeli 9 maddesine yönelik 1000 yanıtlayıcının olduğu veri setinde, kayıp veri içerme durumu, verideki toplam doğru yanlış yanıtları, kayıp veri içeren madde sayıları, en sık ve en az sıklıkla görülen

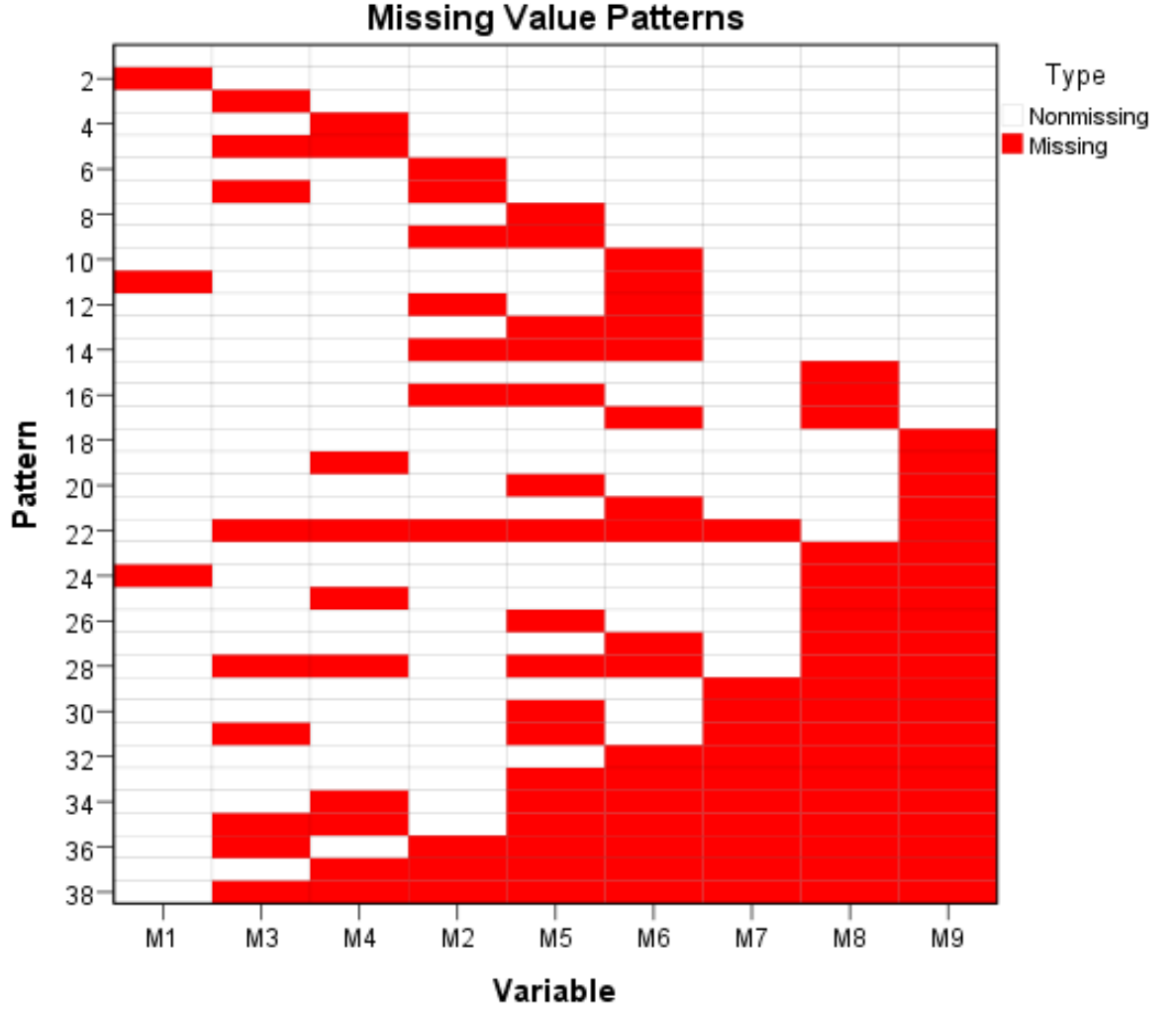
yanıtlar incelenmiştir. Testin uygulandığı yanıtlayıcılardan madde düzeyinde gözlenen ve kayıp verilerin dağılımı Tablo 5’de verilmiştir.

Tablo 5. Kayıp Verilerin Cinsiyete Göre Frekans Dağılımı

	Cinsiyet	m1	m2	m3	m4	m5	m6	m7	m8	m9	
N	Valid	1000	995	974	982	979	939	936	934	916	889
	Missing	0	5	26	18	21	61	64	66	84	111

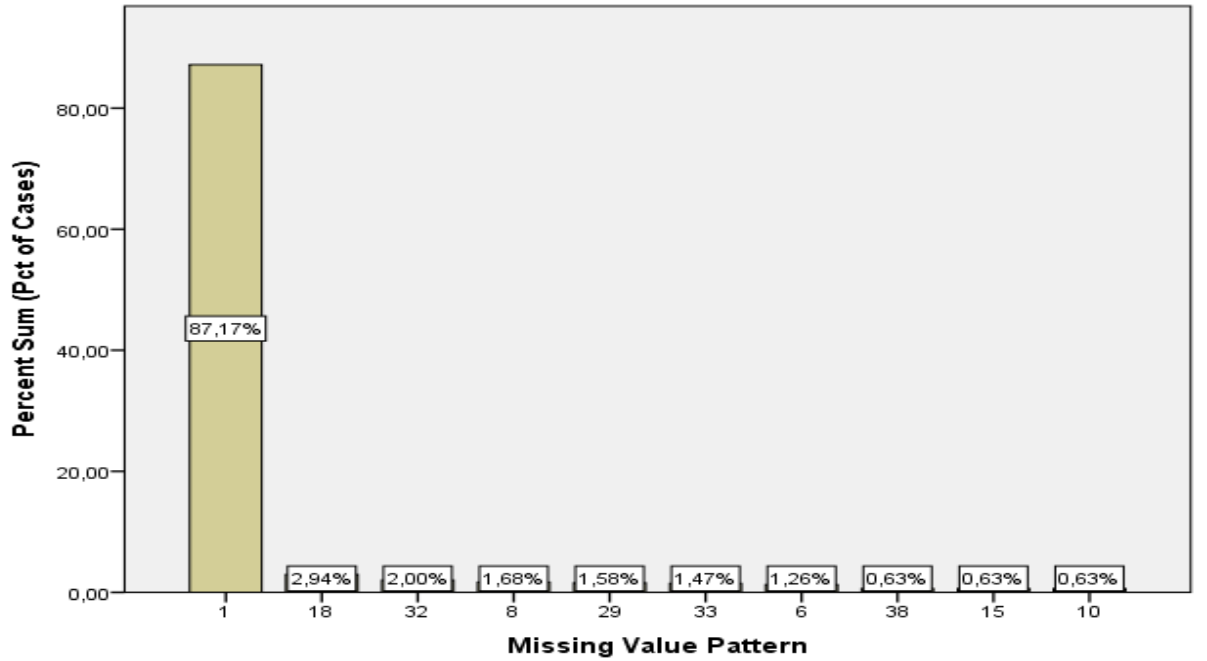
Tablo 5’de cinsiyet değişkeninde kayıp veri bulunmadığı ve uygulanan 9 maddelik (m1-m9) ölçekten elde edilen verilerde her maddeden kayıp veri olduğu gözlenmektedir. Kayıp veriler incelendiğinde kayıp verilerin birinci maddeden dokuzuncu maddeye doğru giderek arttığı gözlenmektedir.

Kayıp veri mekanizmasının daha net şekilde anlaşılması için kayıp veri içeren toplam satır sayısı ve kayıp verilerin veri setine nasıl dağıldığı ve birlikte ortaya çıkma olasılığı incelenmiştir. İncelemeler sonucunda veri setindeki 9 maddenin tamamında kayıpların olduğu, diğer bir ifadeyle her değişkende en az bir kayıp olduğu gözlenmiştir. Elde edilen bulgularda 1000 kişinin 171’inde (%17.1) kayıp veri bulunmadığı, 829’unda (%82.9) ise kayıp veri bulunduğu gözlenmektedir. Yani, veri setindeki 1000 satırdan 829’unda en az bir kayıp veri bulunmaktadır. Ayrıca kayıp verilerin tüm veri setindeki oranı incelenmiştir. Yapılan incelemede 9000 hücreden (9 madde*1000 katılımcı) 8544 hücrede (%94.93) kayıp veri olduğu, 456 hücrede (%5.07) ise kayıp verinin olmadığı gözlenmiştir. Aşağıda kayıpların her bir maddedeki oranına ilişkin Şekil.6 incelendiğinde en fazla kayıp verinin m9’da en az kayıp verinin ise m1’ de olduğu gözlenmektedir. Bu kayıp veri örüntüsüne ait grafik aşağıda Şekil 6’da verilmiştir.



Şekil 6. Kayıp Veri Örüntü Grafiği

Bir veri setindeki kayıp veri örüntüsünün monoton kabul görmesi için kırmızı ve beyaz hücrelerin komşu olması yani yan yana olması gerekir. Yani herhangi bir değişkenden önceki değişkenlerde kayıp olmayacak şekilde sıralanabiliyor ise monoton bir örüntüye sahiptir (Akbaş ve Koğar,2020).Bu bilgiler ışığında kayıp veri örüntüsüne ait grafik incelendiğinde, verilerin monoton olmayan bir görüntü çizdiği gözlenmektedir. Veri setinin monoton olmayan bir veri örüntüsü sergilemekle birlikte, kayıpların değişkenlere dağılımının ilişkili olduğu düşünülmektedir. Yani kayıp verilerin bazı değişkenlerde yoğunlaştığı görülmektedir. En çok kayıp veriye üç değişkende (m7,m8 ve m9) olduğu, kayıpların birlikte ortaya çıktığı görülmektedir. Kayıp veri örüntüsüne ilişkin inceleme sonucunda kayıp veri dağılımına ilişkin frekans dağılımında en sık karşılaşılan 10 örüntünün frekans dağılımı Şekil 7’de verilmiştir.



The 10 most frequently occurring patterns are shown in the chart.

Şekil 7. Kayıp Veri Örüntülerine İlişkin Frekans Dağılımı

Örüntünün frekans dağılımı incelendiğinde, 1 numaralı örüntünün (hiç kayıp veri bulunmayan satırlar) en sık karşılaşılan örüntü (%87.17) olduğu görülmektedir. Diğer örüntülerin frekans dağılımlarının sıralaması ise 18 numaralı örüntü (%2.94 oranla), bu örüntü Tablo.3'te incelendiğinde de bu örüntünün sadece m9 numaralı madde de kayıp veriye sahip bireylerin olduğu görülmektedir. 32 numaralı örüntünün (%2 lik oranla) m6,m7,m8 ve m9 numaralı maddelerde kayıp veriye sahip bireylerin olduğu, 8 numaralı örüntü (%1.68 oranla) m5 madde de, 29 numaralı örüntü (%1.58 lik oranla) m7,m8 ve m9, 33 numaralı örüntü (%1.47 oranla) m5, m6,m7,m8 ve m9, 6 numaralı örüntü (%1.26 oranla) m3, m4, 38 numaralı örüntü (%0.63 oranla) m1 hariç diğer tüm maddelerde, 15 numaralı örüntü (%0.63 oranla) sadece m8, 10 numaralı örüntü (%0.63 oranla) ise m6 değişkeninde görülmektedir.

Kayıp verilerin cinsiyet yani kategorik değişkene göre dağılımı Tablo 6'da verilmiştir. Bu analizde kayıp veri içeren değişkenlerin değerleri "0" kayıp verileri ve "1" gözlenen değerleri temsil etmek üzere yeniden kodlanmıştır.

Tablo 6. Kayıp verilerin Cinsiyet'e Göre Dağılımı

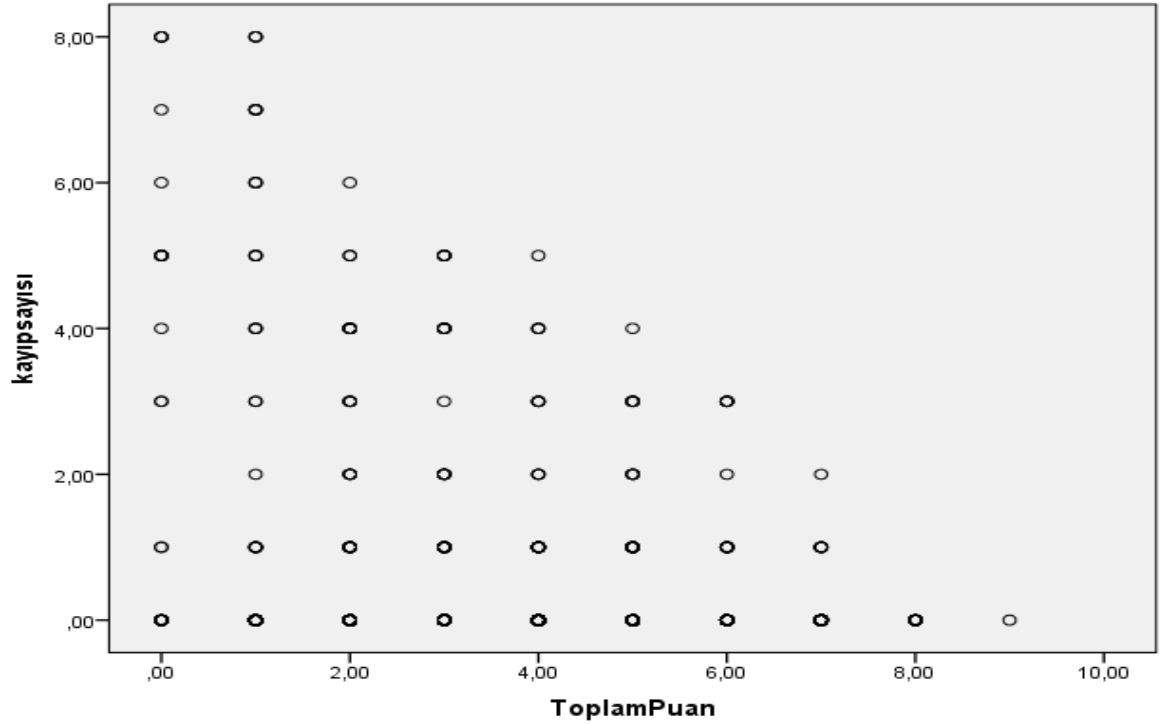
Maddeler	Yanıtlar	Kız	Erkek	Toplam
Km1	.00	2	3	5
	1.00	503	492	995
Km2	.00	9	17	26
	1.00	496	478	974
Km3	.00	10	8	18
	1.00	495	487	982
Km4	.00	9	12	21
	1.00	505	495	979
Km5	.00	25	36	61
	1.00	480	459	939
Km6	.00	25	39	64
	1.00	480	456	936
Km7	.00	27	39	66
	1.00	478	456	934
Km8	.00	33	51	84
	1.00	472	444	916
Km9	.00	42	69	111
	1.00	463	426	889

Kayıp verilerin cinsiyete göre nasıl bir dağılım gösterdiği çapraz tablo incelendiğinde kayıp verilerin yanıtlayıcılara bağlı olarak ortaya çıkmadığı gözlenmektedir. Çapraz tabloda kızlar ile erkeklerin madde yanıtlamada herhangi bir örüntü göstermediği görülmektedir. Yanıtlayıcıların maddelere verdikleri cevapların ortalamalarını bularak maddelerin güçlüklerine göre gözlenen değerler ve kayıp veriler Tablo 7'de verilmiştir.

Tablo 7. Kayıp Sayısına Göre Madde Ortalamalarının Karşılaştırılması

	m1	m2	m3	m4	m5	m6	m7	m8	m9
N Geçerli	995	974	982	979	939	936	934	916	889
Kayıp	5	26	18	21	61	64	66	84	111
Ortalama	0,71	0,38	0,39	0,59	0,43	0,64	0,46	0,21	0,14

Tablo 7’de bulunan maddeler incelendiğinde genellikle maddenin zorluğuna göre kayıp veri sayısının arttığı görülmektedir. Maddelerden en zor maddenin 9. Madde ($p=.14$) olduğu en fazla kaybın (111 kayıp) bu madde de olduğu görülmektedir. En kolay maddenin ise 1. Madde ($p=.71$) olduğu ve bu maddede en az kaybın (5 kayıp) olduğu görülmektedir. Diğer maddelerden m8. maddenin güçlüğü (.21) kayıp veri sayısı 84, m2 maddenin güçlüğü (.38) kayıp sayısı 26 dır. Şekil 8’de ise kayıp sayısı ile toplam puan arasındaki ilişki incelenmiştir.



Şekil 8. Toplam Puan İle Kayıp Sayısı Arasındaki Saçılma Grafiği

Şekil 8’deki grafik incelendiğinde toplam puanı az olanın kayıp sayısının fazla olduğu toplam puan arttıkça kayıp sayısının azaldığı görülmektedir. Buradan hareketle toplam puan ile kayıp sayısı arasındaki korelasyon incelenmiştir. Toplam puan ile kayıp sayısı arasındaki korelasyon (-0.27) olarak elde edilmiştir. Buradan daha başarılı yanıtlayıcıların kayıp sayısının azaldığı, başarısız yanıtlayıcıların ise kayıp sayısının arttığı söylenebilir. Yapılan analizlerin incelenmesi sonucunda elde verilerden hareketle ABİDE-2016 matematik testinin çoktan seçmeli 9 maddesindeki kayıpların SOK’un en net göstergesidir diyebiliriz. Kayıp verilerin meydana gelmesi değişkenin kendisinden kaynaklıdır.

4.2. Test ve Madde Parametreleri

ABİDE-2016 Matematik Testinin çoktan seçmeli dokuz maddesinin dört farklı kayıp yöntemine göre elde edilen eksiksiz verilerinin, KTK temelinde madde parametreleri bağlamında madde güçlüğü, madde ayırt ediciliği dikkate alınmış, test parametreleri bağlamında ise Cronbach alfa güvenilirlik katsayısı, açıklanan varyans, standart sapma, minimum-maksimum değerler, basıklık ve çarpıklık değerlerinin yanında faktör yükleri dikkate alınmıştır. Kayıp veri yöntemlerine göre madde ve test parametreleri Tablo 8 de verilmiştir.



Tablo 8. Kayıp Veri Yöntemlerine Göre Madde ve Test parametreleri

	LW			SRA			HDA			ÇDA		
	p	r	fy	p	r	fy	p	r	fy	p	r	fy
Maddeler												
m1	.70	.33	.59	.71	.30	.62	.71	.32	.59	.71	.30	.62
m2	.37	.29	.50	.37	.26	.50	.38	.29	.50	.38	.28	.52
m3	.39	.24	.40	.40	.22	.40	.39	.26	.40	.40	.23	.41
Madde												
m4	.60	.33	.56	.60	.27	.53	.59	.33	.56	.59	.28	.55
Parametreleri												
m5	.43	.28	.44	.43	.19	.38	.43	.29	.44	.43	.22	.40
m6	.63	.27	.48	.64	.26	.48	.64	.26	.48	.64	.28	.50
m7	.46	.30	.54	.46	.30	.54	.46	.30	.54	.46	.32	.56
m8	.39	.20	.31	.21	.03	.07	.39	.21	.31	.22	.04	.08
m9	.36	.27	.39	.25	-.12	.25	.35	.26	.39	.14	-.11	.23
Güvenirlilik		.58			.47			.59			.49	
Bartlett		1015.0			910.2			1015.0			811.2	
KMO		0.79			0.80			0.79			0.81	
Özdeğer		2.78			2.55			2.83			2.64	
Test												
Avaryans(%)		30.93			28.29			31.52			29.32	
Parametreleri												
Min.		0.00			0.00			0.00			0.00	
Maks.		9.00			9.00			9.00			9.00	
Ortalama		4.32			3.97			4.32			3.96	
S.Sapma		2.1			1.82			2.1			1.85	
Basıklık		-.53			-.65			-.53			.11	
Çarpıklık		.285			.10			.285			-.68	

Tablo 8’de farklı kayıp veri yöntemlerine göre eksiksiz veri setleri üzerinden gerçekleştirilen AFA çıktılarına göre dört kayıp veri yöntemine göre de veri setinin tek boyutlu olduğu görülmüştür. Yapılan çıktılar sonucunda Kaiser-Mayer-Olkin Örneklem Büyüklüğü Uygunluk Testi (KMO) ve Bartlett Küresellik Testi incelenmiştir. Büyüköztürk (2007), KMO değerinin veri matrisinin faktör analizi için uygun olup olmadığı ve veri yapısının faktör çıkarmaya uygunluğu hakkında bilgi vermektedir. KMO değerinin .60 üzerinde olması beklenir. Bartlett testi ise, veri setindeki değişkenler arasında ilişki olup olmadığı hakkında bilgi vermektedir.

Bu bilgiler ışığında tüm kayıp veri yöntemlerinde KMO değeri .60 ve üzerinde olduğu veri setlerinin büyüklüğünün faktör analizi için uygun olduğu görülmüştür. Bartlett Testi kestirimlerine yönelik olasılık değerleri $p < .05$ olarak elde edilmiştir. Tüm kayıp veri yöntemlerine göre elde edilen faktör analizi sonuçlarının tek boyutlu olduğu yani maddeler birinci faktörde toplanmaktadır. Kayıp veri yöntemlerine göre açıklanan varyans oranları %28 ile %32 arasında değiştiği gözlenmektedir. En yüksek varyansın sırasıyla (% 31.52) ile Hot Deck Atama, (% 30.93) ile Listwise, (% 29.32) Çoklu Değer Atama’nın izlediği en düşük açıklanan varyansın ise (% 28.29) Stokastik Regresyon Atamasın’dan elde edilmektedir.

Kayıp veri yöntemlerine göre faktör öz değerleri açıklanan varyans ile doğru orantılı olacak şekilde HDA yönteminde 2.83 ile en yüksek değer, LW yönteminde 2.78, ÇDA yönteminde 2.64 ve en düşük faktör öz değerine ise 2.55 ile SRA yönteminde kestirilmektedir.

AFA uygulamasında faktör yük değerinin .45 veya üzerinde olması seçim için uygun bir ölçüt olmakla birlikte az sayıda maddeler için bu sınır değerin .30’a kadar indirilebilir (Büyüköztürk,2007). Tablo.11 incelendiğinde LW ve HDA yönteminde faktör yük değerleri .31 ile .59 arasında değer almakta, SRA atama yönteminde .07 ile .62 arasında, ÇDA yönteminde ise .08 ile .62 arasında değer almaktadır. LW ve HDA yöntemlerinde tüm maddeler .30 üzerinde olmakla beraber çoğu madde (3,8 ve 9. madde dışında) .45 e yakın veya üzerindedir. SRA ve ÇDA yöntemlerinde .30’un altın iki madde (8. ve 9. madde) bulunmaktadır. Diğer maddeler .40’a yakın ve üzerinde olduğu görülmektedir. Bu veriler ışığında LW ve HDA yöntemlerinin birbirine yakın sonuçlar verdiği, SRA ve ÇDA yöntemlerinin ise birbirine yakın sonuçlar verdiği görülmektedir.

Madde parametreleri bağlamında madde güçlüklerine yönelik Tablo.11 incelendiğinde kayıp veri yöntemlerine göre elde edilen madde güçlük indeksleri arasında sayısal olarak bazı farklılıklar bulunmaktadır. Genel olarak elde edilen madde güçlük indeksleri birbirine yakın ve bazı maddelerde farklılık göstermekle beraber LW yöntemi ve HDA yönteminde daha yüksek, en düşük ise madde güçlük indeksi ise ÇDA elde edilmiştir. Bu kestirimlerle LW ve HDA yöntemlerinin kullanılması ile maddelerin “kolay madde”, ÇDA nın kullanılması ile maddelerin daha “zor madde” değerlendirilmesi olasılığı yüksektir. Madde güçlük ölçütü .20 ile .80 arasında olma ölçütü dikkate alındığında bu ölçütü tüm maddelerde karşılayan LW ve HDA olduğu görülmektedir. Bütün kayıp veri yöntemlerinde güçlüğü .80 üzerinde bir maddeye rastlanmamıştır.

Bir bütün olarak tüm maddeler incelendiğinde en yüksek madde güçlük indeksine sahip yani en kolay maddenin bütün kayıp veri yöntemlerinde 1. Madde (.70 - .71 aralığında) olduğu görülmektedir. En düşük madde güçlük indekslerine yani en zor maddenin LW yönteminde ve HDA yönteminde 1. Ve 9. madde, SRA ve ÇDA yöntemlerinde ise 8. Ve 9. maddelerdir. Buradan tüm yöntemlerde en düşük madde güçlük indeksine sahip yani en zor madde 9. Maddedir.

Diğer bir madde parametresi olan madde ayırt edicilik indeksi kestirimleri yapılmıştır. Madde ayırt edicilik indeksi, ölçülmek istenen özelliğe sahip yanıtlayıcılar ile bu özelliğe sahip olmayan yanıtlayıcıları ayırt etmesi gerekmektedir. Madde ayırt ediciliği .20 ve üzeri değerlerin genel kabul gören bir değer görülmektedir. (Lord ve Novick, 1968; Tekin, 2017 ve Baykul, 2000).

Madde parametreleri bağlamında Tablo.11 incelendiğinde kayıp veri yöntemlerine göre madde ayırt edicilik indeksleri arasında sayısal farklılıklar gözlenmektedir. Genel olarak LW ve HDA atama yöntemlerinde birbirine yakın olmakla beraber daha yüksek madde ayırt edicilik indeksleri görülmektedir. En düşük ayırt edicilik indekslerine SRA yönteminde görülmektedir.

Madde ayırt edicilik indeksleri LW yönteminde .20 ile .33 arasında, HDA yönteminde .21 ile .32 arasında değişmekte, SRA yönteminde .03 ile .30 arasında ve ÇDA yönteminde .04 ile .30 arasında değişmektedir. En geniş aralığın SRA atama yönteminde olduğu görülmektedir.

Tüm yöntemler incelendiğinde en yüksek ayırt edicilik indeksine sahip maddenin 1. madde olduğu ve en düşük ayırt edicilik indeksine sahip maddenin ise 8. Madde olduğu görülmektedir. SRA ve ÇDA yöntemlerinde.20 altında 8. ve 9. maddeler bulunmaktadır.

Madde güvenilirlik katsayısı bu çalışmada dikkate alınan bir diğer parametresidir. Güvenirlik bir ölçme aracının doğruluğu veya duyarlılığı anlamına gelir. Güvenirlik indeksleri, belli bir ölçme aracından elde edilen puanların ne derece tutarlı ve tekrarlanabilir olduğu hakkında bir fikir verir. Bir testin neyi ölçtüğünü değil, ölçtüğü şeyi ne kadar doğru ölçtüğünü belirler (Thorndike, 1982). İki kategorili puanlanan maddelerde genellikle puan aralığının dar olması sebebiyle çok anlamlı görülmemektedir. İki kategorili puanlanan maddelerde toplam puana yönelik güvenilirlik kestirimlerinin yapılması daha doğru bilgilere ulaşmaya yardımcı olacaktır. (Baykul,2000; Atılğan, 2009).

Tablo 8, incelendiğinde bütün kayıp veri yöntemlerinde de güvenilirlik katsayısının düşük olduğu görülmektedir. En yüksek güvenilirlik katsayısına HDA yönteminde .59 ile ulaşılmakta bu yöntemi .58 LW yöntemi izlemekte, en düşük güvenilirlik katsayısına ise .47 ile SRA yönteminde görülmektedir.

Kayıp veri yöntemleri incelendiğinde veri yayılma ölçülerinde biri olan standart sapmanın LW ve HDA yöntemlerinde benzer bir sonuç ile 2.1 olarak çıkmaktadır. SRA ve ÇDA yöntemlerinde standart sapma değerleri birbirine yakın olmakla birlikte en düşük standart sapma değerine SRA da ulaşılmaktadır.

Kayıp veri yöntemlerine göre ortalama puanlar incelendiğinde LW ve HDA atamanın diğer verilerde olduğu gibi yine benzer bir sonuçla karşılaşılmakta bu iki yöntemin ortalamaları 4.32 olarak çıkmaktadır. En düşük ortalamaya ise ÇDA yönteminde 3.96 ile ulaşılmaktadır.

Tablo 8, incelendiğinde basıklık ve çarpıklık durumları incelendiğinde LW, HDA ve SRA yöntemlerinin sağa çarpık bir dağılıma sahip olduğu bu üç yöntemden farklı olarak ÇDA yönteminin sola çarpık bir dağılım sergilediği görülmektedir.

Demir (2013), yılında elde ettiği parametre sonuçları bu yapılan çalışmayı destekler nitelikte silme yöntemlerinin test güvenilirliğini arttırdığını söylemektedir. Yapılan bu çalışmada da LW yönteminin HDA atama ile benzerlik gösterdiği fakat

diğer iki atama yöntemine göre daha yüksek kestirimler elde edildiđi görölmektedir. Demir (2013), Seçkisiz olmayan kayıp verilerin bulunduđu iki kategorili veri setlerinde uygun kayıp veri yönteminin belirlenmesini ifade etmiştir. Bu çalışmada bu ifadeyi destekler nitelikte yapılan kestirimler de seçkisiz olmayan kayıp veri yöntemlerinde LW ve HDA atama yönteminin daha verimli sonuçlar verdiği gözlenmiştir.

Yapılan diğer çalışmalar incelendiğinde örneklem büyüklükleri, veri örüntüleri ve mekanizmaları incelendiğinde silme yöntemlerinin diğer yöntemlere göre düşük kestirimler elde edilmiştir. Fakat yapılan bu çalışmada kayıp veri mekanizması seçkisiz olmayan kayıp olduđu zaman kestirimlerin daha yüksek sonuçlar verdiği gözlenmiştir. Bu durum iki kategorili veri setlerinde kayıp veri mekanizmasının, seçkisiz olmayan kayıp veri mekanizmasına sahip olması durumun da uygun kayıp veri yöntemleri daha dikkatli incelenmelidir.

BEŞİNCİ BÖLÜM

SONUÇ ve ÖNERİLER

5.1. Sonuçlar

ABİDE-2016 verileri incelendiğinde kayıp verilerin herhangi bir düzen içinde, planlı bir şekilde veya monoton bir şekilde olmadığı anlaşılmaktadır.

Testi alan bireylerden 171 kişide (%17.1) kayıp veri bulunmamakta yani doğru veya yanlış olacak şekilde bir cevap vermiştir. Diğer yandan 829 kişide (%82.9) ise en az bir madde de kayıp veri bulunmaktadır. Testi alan yanıtlayıcıların %50.5 i erkeklerden %49.5 i kızlardan oluşmaktadır. Cinsiyet değişkeni dikkate alındığında cinsiyete göre bir farklılık olmadığı gözlenmemiştir.

Maddeler düzeyinde kayıp veri incelendiğinde en çok kayıp 111 kayıp ile 9. Madde, en az kayıp ise 5 kayıp ile 1. maddedir. Maddeler düzeyinde kayıplar incelendiğinde kayıpların birinci maddeden başlayarak dokuzuncu maddeye doğru doğrusal olmasa da bir artış gösterdiği gözlenmektedir. Toplam puan ile kayıp sayısı arasındaki korelasyon -.27 bulunmuştur. Toplam puan arttıkça kayıp sayısının düştüğü gözlenmiştir.

Kayıp veri mekanizmasına yönelik analiz ve incelemelerde ‘Little MCAR Test’ kestirimleri yapılmıştır. Yapılan ‘Little MCAR Test’ kestirimleri ($\chi^2 = 237.164$, $P > .05$), kayıpların TSK mekanizmasına sahip olduğunu göstermektedir. ABİDE-2016 Matematik A kitapçığındaki 9 madde incelenmiştir. Her bir madde düzeyinde kayıp veri bulunması, her madde de 1 ve 0 puanlamaların bulunması, kayıp verilerin genel ve dağınık örüntü göstermesi, yanıtlayıcılar ile verdikleri yanıt örüntüleri arasında benzerlik olmaması, cevapların ve kayıp verilerin cinsiyet ile arasında benzerlik göstermemesi TSK varsayımını karşılada SOK varsayımını karşılama olasılığının yüksek olduğu belirlenmiştir.

Kayıp veriler veri setleri için istenen bir durum olmamakla birlikte yapı geçerliliğini olumsuz yönde etkilemektedir. Kayıp veriler ile baş etmede en iyi yaklaşım hiç kaybın olmamasına yönelik çalışma sarf edilmelidir. Kayıp verilerin ihmal edilemeyecek düzeyde fazla olması yapı geçerliliği için yapılacak analizlerde olumsuz sonuçlar elde edileceğinden, öncelikle kayıp veri yöntemlerine göre kayıp veri ataması yapılmalıdır. Kayıp verilerin atanması sonucunda eldeki veriler ile Açıklayıcı Faktör Analizi (AFA) incelenmiştir.

ABİDE-2016 Matematik Testi A kitapçığının çoktan seçmeli 9 maddesinin yapı geçerliliğine yönelik olarak veri seti üzerinden yapılan AFA sonuçlarına göre atama yapılan bütün kayıp veri yöntemlerinde veri setinin tek faktörlü olduğu görülmüştür. Kayıp veri yöntemlerinin yapı geçerliliğine yönelik AFA sonuçları karşılaştırıldığında LW ve HDA atamanın benzer sonuçlar, SRA ve ÇDA yöntemlerinin ise birbirine daha yakın sonuçlar verdiği gözlenmektedir.

Madde ve test parametreleri klasik test kuramı bağlamında incelenmiştir. Madde güçlük indeksi kestirimleri bütün kayıp veri yöntemlerin de .80 ve altında değer almaktadır. Bütün yöntemlerde madde güçlük indeksi en yüksek yani en kolay olan maddenin 1. madde olduğu görülmüştür. Madde güçlük indeksi genel olarak LW ve HDA atama yöntemlerin de .20 altında maddelere rastlanmamıştır. Diğer iki yöntem olan SRA’da 8. ve 9. madde .20’nin altında, ÇDA yönteminde ise 8. Madde .22 ve 9. madde .14 değerlerini almıştır. Bu bilgiler ışığında LW ve HDA yöntemlerinde 1 ile puanlanan madde sayılarının arttığı, SRA ve ÇDA atama yöntemlerinde 0 ile puanlanan madde sayılarının arttığı gözlenmektedir. Buradan bu yöntemlerin yanlılığa yol açma olasılığının yüksek olduğu düşünülmektedir. SRA ve ÇDA yöntemlerinde

madde güçlük indekslerinin düşüşe yol açtığı görülmektedir. LW ve HDA atama yöntemlerinde madde güçlüklerinin birbirine daha yakın sonuçlar verdiği, SRA ve ÇDA atama yöntemlerinde ise açıklığın daha fazla olduğu görülmektedir.

Madde ayırt edicilik indeksleri dikkate alındığında LW ve HDA atama yöntemlerinde bütün maddeler .20 ve üzerinde değer alırken SRA ve ÇDA atama yöntemlerinde .20 altında değerler elde edilmektedir. SRA yönteminde 5 ,8 ve 9. Maddeler, ÇDA yönteminde 8. Ve 9. maddeler .20 altında değer aldığı için yeniden gözden geçirilebilirler. Buradan hareketle LW ve HDA atama yöntemleri atanan kayıp verilerle maddelerin daha yüksek ayırt edicilik indeksi ürettiği gözlemlenmiştir.

5.2. Öneriler

İleride yapılacak araştırmalara yönelik olarak;

1. Benzer incelemelerin ABİDE-2016'da yer alan Matematik alt testinin dışında diğer alt testler içinde gerçekleştirilmesi,
2. ABİDE-2016 sınavının psikometrik özelliklerinin bu çalışmada kullanılan LW, HDA, SRA ve ÇDA dışında daha fazla veri kayıp veri yöntemlerine göre incelenmesi,
3. Seçkisiz olmayan kayıp verilerin varlığında kayıp veri atama yönteminin seçiminde uygun yöntemin seçimine dikkat edilmesi gerekmektedir.
4. Bu çalışmada kullanılan kayıp veri yöntemlerine göre testin psikometrik özelliklerinin Madde Tepki Kuramı temelinde incelenmesi önerilmektedir.

KAYNAKÇA

Acock, A.A. (2005). Working with missing values. *Journal of marriage and family*, 65(4), 1012-1028

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

Akbaş, U. (2014). *Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi*. Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Akbaş, U. ve Koğar, H. (2020). Nicel araştırmalarda kayıp veriler ve uç değerler: *Çözüm önerileri ve SPSS uygulamaları*. (1. Baskı). Ankara: Pegem Akademi.

Akbaş, U. ve Tavşancıl, E. (2015). Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 38-57. doi.org/10.21031/epod.26476.

Akyıldız, M. ve Karadağ, N. (2018). Farklı soru türlerinin ve ayırt edicilik düzeylerinin incelenmesi. *Açık Öğretim Uygulamaları ve Araştırmaları Dergisi*.4 sayı(1) 112-122.

Allison, P. D. (2002). *Missing data*. California: Sage Publication, Inc.

Allison, P. (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. Ed. Roger E. Millsao ve Alberto Maydeu-Olivares, *Quantitative Methods in Psychology*, sy.72-89. London: SAGE Publication

Alpar, R. (2003). *Uygulamalı çok değişkenli istatistiksel yöntemlere giriş-1*, Nobel Kitabevi.

Atılğan, H. (2017). Türkiye’de kademeler arası geçiş: dünü-bugünü ve bir model önerisi. *Ege Eğitim Dergisi*, 19(1), 1-18.

Arıcı, H.(1991). *İstatistik yöntem ve uygulamalar*. Ankara: Meteksan A.Ş.

Bahar, M., Nartgün, Z., Durmuş, S. ve Bıçak, B. (2012). *Geleneksel tamamlayıcı ölçme değerlendirme teknikleri (5. Baskı)*. Ankara: Pegem Akademi.

Baki, A. ve Birgin, O. (2002). *Matematik eğitiminde alternatif bir değerlendirme olarak bireysel gelişim dosyası uygulaması*. V. Ulusal fen bilimleri ve matematik eğitimi kongresi bildiri kitabı, II, 913-920. Ankara: Devlet Kitapları Müdürlüğü.

Baraldi, A.N. & Enders, C.K. (2010). An Introduction to Modern Missing Data Analysis. *Journal of School Psychology*, 48, 5–37

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları

Büyüköztürk, Ş. (2007). *Sosyal bilimler için veri analizi el kitabı*. 8. Baskı, Ankara: Pegem Akademi.

Büyüköztürk, Ş. (2015). *Sosyal bilimler için veri analiz el kitabı: İstatistik, araştırma deseni SPSS uygulamaları ve yorum*, (Genişletilmiş 21. Baskı). Ankara: Pegem Akademi.

Camilli, G., ve Shepard, L. A. (1994). *Methods for identifying biased test items*. California: Sage.

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, C-Y. (2011). A Simulation Study Using EFA and CFA Programs Based on the Impact of Missing Data on Test Dimensionality, *Expert Systems With Applications*. 39(2012), 4026 – 4031.

Chumney, F.(2012). Measurement in Educational Research. <https://pdfs.semanticscholar.org/d614/09f49f9a4f8290376b9eecf7dafcc4a40a6f.pdf>

Clauser, B. E., ve Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44.

Courville, T.G. (2004). *An Empirical Comparison of Item Response Theory and Classical Test Theory Item/Person Statistic*. Yayınlanmamış doktora tezi, Texas A&M University. Educational and Psychological Measurement.

Crocker, L., and Algina, J. (1986). *Introduction Classical and Modern Test Theory*. USA: CBS College Publishing Company

Çokluk, Ö., ve Kayrı, M. (2011). *Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenilirliği üzerindeki etkisi*. Kuram ve Uygulamada Eğitim Bilimleri (11), 289-309.

De Luca, G., ve Peracchi, F. (2007). A Sample Selection Model For Unit and Item Nonresponse in Cross-sectional Surveys. *CEIS Tor Vergata - Research Paper Series*. 33(99).

Demir, E. (2013). *Kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin incelenmesi*. Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Demirtaşlı, N. (2010). Üst düzey düşünme becerilerinin ölçülmesinde gündelik yaşam unsuru. *Cito Eğitim: Kuram ve Uygulama*. 7, 9-26.

Ebel, R. L. and Frisbie, D. A. (1991). *Essentials of Educational Measurement* (5. baskı). New Delhi: Prentice-Hall.

Education, Audiovisual and Culture Executive Agency. (2009). National testing of pupils in Europe: Objectives, organisation and use of results. Brussels: EURYDICE. Erişim adresi: /Downloads/national_testing_of_pupils_in_europe_090928.pdf

Ekinci, O. ve Bal, Ayten P. (2019). 2018 Yılı Liseye Geçiş Sınavı (LGS) Matematik Sorularının Öğrenme Alanları ve Yenilenmiş Bloom Taksonomisi Bağlamında Değerlendirilmesi. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi* 2019 7(3) 9–18.

Embretson, S.E. ve Reise, S.P. (2000). *Item Response Theory for Psychologists*. Mahway.

Enders, C. K. (2003). Using the Expectation Maximization Algorithm to Estimate Coefficient Alpha for Scales With Item-Level Missing Data. *Psychological Methods*. 8(3), 322-337.

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.

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

Erkuş, A. (2003). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi Yayınları.

Ersoy, Y. (1997). Okullarda Matematik Eğitimi: Matematikte Okuryazarlık, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, sayı:13, 115-120.

Fan, X. (1998). Item Response Theory and Classical Theory: An Empirical Comparison of Their Item-Person Statistic. *Educational and Psychological Measurement*, v:58 n:3, 357-381.

Ferguson, G.A. (1976) *Statistical Analysis In Psychology and Education*. New York: McGraw-Hill.

Francisco, J.S., Nakhleh, M.B., Nurrenbern, S.C. ve Miller, M.L. (2002). Assessing student understanding of general chemistry with concept mapping. *Journal of Chemical Education*, 79(2), 248-257.

Garrett, P. (2009). *A monte carlo study investigating missing data, differential item functioning and effect size* (Doktora tezi). Georgia State Üniversitesi, Atlanta.

Graham, J. W., (2012). *Missing Data: Analysis and Design*. New York: Springer.

Gronlund, N.E. ve Linn, R.L. (1990). *Measurement and Evaluation in Teaching*. McMillan Company, New York.

Haladyna, T. M. (1997). *Writing Test Item to Evaluate Higher Order Thinking*. USA: Allyn ve Bacon.

Hambelton, R.K. (1994). Item Response theory: A broad psychometric framework for measurement advances. *Psicothema*, 6, 3, 535-556.

Hambleton, R.K. ve Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*.12(3), 38-47.

Hartley, H.O., Hocking, R.R. (1971): The Analysis of Incomplete Data, in: *Biometrics*, 27, S.783-823

Heerwegh, D. (2005). *Web Surveys. Explaining and Reducing Unit Nonresponse, Item Nonresponse And Partial Nonresponse*. Yayınlanmamış doktora tezi, Katholieke Universiteit Leuven Faculteit Sociale Wetenschappen, Leuven, Belçika.

Hohensinn, C. ve Kubinger K. D. (2011). On the impact of missing values on item fit and the model validness of the Rasch model. *Psychological Test and Assessment Modeling*,53,380-393

Howell, C.D. (1992) *Statistical Methods For Psychology*. Third Edition, California: Duxbury Press.

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.

Huba, M. E. ve Freed, J.E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning*. Needham Heights, MA: Allyn & Bacon.

Kan, A. (2017). Ölçmenin Temel Kavramları. ,H. Atılğan, (Ed.). *Eğitimde Ölçme ve Değerlendirme* (10. Baskı). Ankara: Anı Yayıncılık.

Kizlik, B. (2012). *Measurement, assessment, and Evaluation in Education*.

Köse, 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.

Kutlu, Ö. Doğan, C. D. ve Karakaya, İ. (2014). *Ölçme ve değerlendirme: Performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi Yayıncılık.

Küçükahmet, L. (2004). *Öğretimde Planlama ve Değerlendirme*. Ankara: Nobel Yayın Dağıtım.

Little, RJA. (1998). A test of missing completely at random for multivariate veri with missing values. *Journal of the American Statistical Association* 38, 11981202

Little, R. J. ve Rubin, D. B. (1987). *Statistical Analysis With Missing Data* (2nd ed) *New York: Wiley*

Lord, F.M., 1980. *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, 274s, Hillsdale, NJ.

Lord, F.M., Novick. M.R. (1968). *Statistical Theories of Mental Test Scores*. Addison Wesley Publishing Company. Educational Testing Service.

Macdonald, P., Paunonen S.V. (2002). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory. *Educational and Psychological Measurement*, 62. 921-943.

MEB, (2005). *PISA 2003 Projesi Ulusal Nihai Raporu*. Ankara: Milli Eğitim Basımevi.

MEB (2009). İlköğretim öğrencilerinin başarılarının belirlenmesi (ÖBBS): Durum belirleme raporu. Eğitim Araştırma Geliştirme Dairesi (EARGED), Ankara: MEB yayınları.

MEB (2010). Seviye belirleme sınavının değerlendirilmesi raporu. Eğitim Araştırma Geliştirme Dairesi (EARGED), Ankara: MEB yayınları.

MEB (2014). Ortaöğretim Kurumlarına Geçiş Yönergesi Tebliğler Dergisi : Ocak 2014/2676 Erişim Adresi: <http://mevzuat.meb.gov.tr/dosyalar/1679.pdf>

MEB (2015). Strateji Geliştirme Başkanlığı 2015–2019 Stratejik Planı. Ankara

MEB (2016a). Uluslararası öğrenci değerlendirme programı PISA 2015 ulusal raporu.

https://odsgm.meb.gov.tr/test/analizler/docs/PISA/PISA2015_Ulusal_Rapor.pdf

sayfasından erişilmiştir.

MEB (2016b). TIMSS 2015 Ulusal matematik ve fen bilimleri ön raporu 4. ve 8. sınıflar. <https://docplayer.biz.tr/130870560-Timss-uluslararasi-matematik-ve-fen-egilimleri-arastirmasi-timss-2015-ulusal-matematik-ve-fen-bilimleri-on-raporu-4-ve-8-siniflar.html> sayfasından erişilmiştir.

MEB (2017c). Akademik Becerilerin İzlenmesi ve Değerlendirilmesi-8.Sınıflar Raporu. Ankara: ÖDSGM

https://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_11/30114819_iY-web-v6.pdf

sayfasından erişilmiştir.

- MEB (2019). Strateji Geliştirme Başkanlığı 2019–2023 Stratejik Planı. Ankara
- Molenberghs, G., Kenward, M.G. (2007). Missing Data in Clinical Studies. *John Wiley and Sons, Chichester, UK.*
- Nartgün, Z. (2002). *Aynı tutumu ölçmeye yönelik Likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin klasik test kuramı ve örtük özellikler kuramına göre incelenmesi.* Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation: An International Journal of Theory and Practice*, 7 (4), 353-383. doi.org/10.1076/edre.7.4.353.8937.
- Rubin, D. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley ve Sons.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* New York: Chapman & Hall.
- Schafer, J. L., Maren K., ve Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33 (4), 545-571.
- Schafer, J. L. ve Graham, J. W. (2002). Missing Data: Our View Of The State Of The Art. *Psychological Methods*.
- Stage, C. (2003). Classical Test Theory or Item Response Theory: The Swedish Experience. *Em No: 42, Repon, Sweden Umea University. Department of Educational Measurement.*
- Stiggins, R.J.(1999). Assessment, student confidence and school success. *Phi Delta Kappan*, 81.
- Şahin Kürşad, M. ve Nartgün, Z. (2015). Kayıp veri sorununun çözümünde kullanılan farklı yöntemlerin ölçeklerin geçerlik ve güvenirliği bağlamında karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi* , 6(2), 254-267. doi: 10.21031/epod.95917
- Şişman, M. Acat, M.B. Aypay, A. ve Karadağ, E. (2011). *TIMSS 2007 ulusal matematik ve fen raporu 8. sınıflar.* Ankara: Milli Eğitim Bakanlığı.

Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics*. (4.th ed). Boston: Ally ve Bacon.

Tan, Ş. ve Erdoğan, A. (2004). *Öğretimi Planlama ve Değerlendirme*. Ankara: Anı Yayıncılık.

Tekin, H. (2017). *Eğitimde ölçme ve değerlendirme* (23. Baskı). Ankara: Yargı.

Thorndike, R L. (1982). *Applied Psychometrics*. Boston: Houghton Mifflin Cornpany.

TTKB, (2005). *İlköğretim 1-5. Sınıf programları tanıtım el kitabı*, Ankara: MEB Talim ve Terbiye Kurulu Başkanlığı.

OECD (2019a). *PISA 2018 Assessment and analytical framework*. Paris: OECD Publishing. doi:<https://doi.org/10.1787/b25efab8-en>

Turgut, M.F. ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Yayınları.

Uebersax, John S. (2015). "Introduction to the Tetrachoric and Polychoric Correlation Coefficients." *Obtenido de Http://www. John-Uebersax. Com/Stat/Tetra. Htm.[Links]*.

Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman & Hall/CRC Press.

