

Kayıp Verilerin Varlığında Çoktan Seçmeli Testlerde Madde ve Test Parametrelerinin Kestirilmesi: SBS Örneği¹

Ergül DEMİR²

ÖZET

Bu çalışmada, kayıp verilerin varlığında, çoktan seçmeli testlerde, farklı kayıp veri yöntemleri kullanılarak kestirilen madde ve test parametreleri arasındaki ilişkilerin incelenmesi ve bu tür testlerde kullanılması uygun olan kayıp veri yöntemlerinin belirlenmesi amaçlanmıştır. Temel araştırma türünde, ilişki tarama modelinde bir araştırma olarak tasarlanan bu çalışmada, analizler 527517 yanıtlayıcıya yönelik SBS (Seviye Belirleme Sınavı) 2011 Matematik Testi A Kitapçığı verileri üzerinde yürütülmüştür. Veri analizlerinde silmeye dayalı yöntemlerden 'dizin silme yöntemi', basit atama yöntemlerinden '0 atama', 'seri ortalamaları ataması', 'gözlem birimi ortalaması ataması', 'yakın noktalar ortalaması ataması', 'yakın noktalar medyan ataması', 'doğrusal interpolasyon' ve 'dorsal eğilim noktası ataması' yöntemleri, en çok olabilirlik yöntemlerinden 'regresyon atama', 'beklenti-maksimizasyon algoritması' ve 'veri çoğaltma' yöntemleri, çoklu veri atama yöntemlerinden ise 'Markov zincirleri Monte Carlo' yöntemi olmak üzere 12 farklı kayıp veri yöntemi kullanılmıştır. Elde edilen bulgular, kayıp verilerin ihmal edilebilir olmaması durumunda çoktan seçmeli testlere yönelik istatistiksel kestirimlerde, uygun bir kayıp veri yönteminin kullanılmasının gerekli olduğunu göstermektedir. Silmeye dayalı yöntemler ve 0 Atama yöntemi, bu tür veriler için uygun yöntemler değildir. Basit atama yöntemlerinin ise yanlı kestirimler üretme olasılığı yüksektir. En çok olabilirlik ve çoklu veri atama yöntemleri, bu tür verilerde kullanılması en uygun kayıp veri yöntemleri olarak değerlendirilmektedir.

Anahtar Kelimeler: Kayıp veri, İhmal edilebilirlik, Seçkisiz kayıp, Kayıp veri yöntemleri, SBS

 DOI Number: <http://dx.doi.org/10.12973/jesr.2013.324a>

¹ Bu çalışma, Prof. Dr. Nizamettin Koç danışmanlığında hazırlanan Doktora tezinden yararlanılarak hazırlanmıştır.

² Öğr. Gör. Dr. - Ankara Üniversitesi Eğitim Bilimleri Fakültesi - erguldemir@ankara.edu.tr

GİRİŞ

Kayıp veri (missing value), amaçlanan ölçmeye yönelik bir gözlemin bulunmaması durumunu tanımlamaktadır. Psikolojik ölçmelerin temeli, gözlenen verilerden hareketle örtük değişkenlere yönelik çıkarımların yapılmasıdır. Gözlenen verilerdeki eksikler, bu tür çıkarımların yapılabilmesi önündeki en önemli engeli oluşturmaktadır (Hohensinn & Kubinger, 2011).

İstatistiksel analizlerde kayıp veri sorunu, önemli bir tartışma alanıdır. Allison (2002) ve Graham'a (2009) göre bu tartışmaların çıkış noktası, araştırmacılar tarafından kullanılan ve çoğu yirminci yüzyılın başlarında geliştirilmiş olan analitik süreçlerin, eksiksiz veri setleri üzerinden yapılandırılmış olmasıdır. Bu analitik yaklaşımların neredeyse tamamı, tam bilgi gerektirmektedir ve kayıp veriler içeren yanıtlara yönelik herhangi bir çözüm mekanizması içermemektedir.

Standart istatistiksel yöntemler, dikdörtgenel veri setlerinin analizine yönelik olarak geliştirilmiştir. Bir matris şeklinde hazırlanan bu veri setlerinde satırlar gözlemleri, sütunlar ise değişkenleri temsil etmektedir. Matrisin 'hücre' olarak adlandırılan her bir birimine gerçek sayı değerleri girilir. Gerçek sayılar yaş, gelir gibi sürekli değişkenleri, eğitim düzeyi gibi sıralı kategorik değişkenleri ya da cinsiyet, uyruk gibi sınıflamalı kategorik değişkenleri temsil eder. Bir değişkene yönelik bir gözlemin bulunmaması durumunda, söz konusu gözlemi temsil eden hücre boş kalır. Temel bir yaklaşım olarak bir yanıtın eksikliği, ölçülen değişkene yönelik örnek uzayda ek bir nokta ile temsil edilir. Bu nokta evrenin bir 'kayıp veri' alt tabakasını oluşturmaktadır. Bir takım özel kodlamalar aracılığı ile bir değişkene yönelik kayıp veri alt tabakasının tanımlanması mümkündür. Bu tür bir tanımlamada gerektiğinde birden fazla kodlama kullanılabilir. Örneğin 'bilgisi olmama', 'yanıtlamayı reddetme' ya da 'yasal sınırların dışında olma' gibi farklı kodlamalar yapılabilmektedir. Kayıp veri, veri matrisine 'gözlenemeyen veri' olarak girilir ve kayıp veriler içeren bir veri seti ortaya çıkar (Little & Rubin, 1987).

Verilerde kayıpların ortaya çıkmasına neden olan en belirgin yol, bazı yanıtlayıcılardan ya da gözlemlerden hiçbir bilgi elde edilememesidir. Örneğin eğitim alanında öğrencilerin bir kısmı, ailelerin onayı olmadığı için ya da başka bir okula nakil gittikleri için test uygulamasına katılmayabilmektedir. Benzer şekilde öğretmenlerden veri toplanan bir araştırmada, öğretmenlerin bir kısmı test almayı reddedebilmekte ya da uygulamanın yapıldığı zaman okulda bulunamayabilmektedir. Bu durum 'birim düzeyinde yanıtlanmama (unit nonresponse)' olarak tanımlanmaktadır. Bir veri toplama aracının tamamına yanıt verilmemesi durumuna ek olarak sıklıkla karşılaşılan bir diğer sorun, yanıtlayıcının bazı maddeleri yanıtlamayı reddetmesi durumunda ortaya çıkmaktadır. Yanıtlayıcı, özellikle zihinsel davranışları ölçen testlerde ya da tipik performans testlerinde, yanıtı bilmeme, madde ile ölçülmek istenen özellik açısından yeterli olmama, maddeyi kazara atlama gibi nedenlerle bazı maddeleri yanıtlamayabilmektedir. Bu durum ise 'madde düzeyinde yanıtlanmama (item nonresponse)' olarak tanımlanmaktadır (Rubin, 1987; Puma, Olsen, Bell & Price, 2009).

Kayıp veri, veri setinin daralması ve buna bağlı olarak yapılacak kestirimlerin gücünün azalması anlamına gelmektedir. Kayıp veriler içeren bir veri seti üzerinde, eksiksiz veri setlerine göre yapılandırılmış standart analiz yöntemlerinin kullanılması da mümkün değildir. Dahası, yanıtlayanlar ile yanıtlamayanlar arasındaki, sıklıkla sistematik olan farklılıklardan dolayı, olası bir yanlılık söz konusudur. Yanıtlamayanların ve yanıtlamama gerekçesinin genellikle bilinmiyor olması da, bu tür bir yanlılığın giderilmesini zorlaştırmaktadır (Rubin, 1987).

Allison'a göre (2002) araştırmacılar genellikle, kayıp verilerin gözlenen ölçmelerde manidar bir farka yol açmadığını kanıtlama eğilimindedir. Örneğin, gelirlerini beyan eden ve etmeyen yanıtlayıcılar arasında, dikkate alınan değişkenler açısından manidar bir fark oluşmadığına yönelik kanıt sağlama çabası oldukça yaygındır. Daha genel bir durum olarak araştırmacılar, ne anlama geldiğini açıkça anlamamalarına rağmen sıklıkla verilerinin 'seçkisiz kayıp (missing at random)' içerdiğini varsaymakta ve bu varsayıma bağlı olarak kayıp verileri analiz dışı bırakmaktadırlar.

Demir ve Parlak (2012), eğitim araştırmalarında kayıp veri sorununu ele aldıkları çalışmalarında araştırmacıların, genellikle eksiksiz veri seti üzerinde çalışma eğiliminde olduklarını belirlemiştir. Kayıp verilerin karakteristiğinin incelendiği araştırma sayısı oldukça azdır. Kayıp veriler, sıklıkla herhangi bir istatistiksel kanıt olmaksızın ihmal edilmekte ve analiz dışı bırakılmaktadır. Benzer bir araştırmada Groves (2006) araştırmacıların, herhangi bir istatistiksel kontrol kullanmaksızın ve sıklıkla herhangi bir rasyonel gerekçe sunmaksızın, değişkenler düzeyinde % 15 ile % 50 arasında kayıp veriyi ihmal edebildiğini belirlemiştir.

Rubin'e (1976) göre çoğu araştırmada kayıp verilerin analiz dışı bırakılmasının temelinde bu verilerin, bir takım varsayımlara bağlı olarak 'ihmal edilebilir (ignorable)' görülmesi yatmaktadır. Örneğin kayıp verilerin seçkisiz olarak oluştuğu, bu nedenle veri dağılımında bir sapma ya da farklılık oluşmayacağı, çok değişkenli normal dağılım varsayımının sağlanması durumunda her bir değişkene yönelik kayıp veri oluşma olasılığının eşit olacağı ya da bağımlı değişkenin, gözlenen değişkenle ilgisi olmayacak şekilde kayıp veri içerebileceği gibi varsayımlara bağlı olarak kayıp verilerin analiz dışı bırakıldığı araştırmalara sıklıkla rastlanmaktadır. Allison'a (2002) göre esas itibarıyla ihmal edilebilirlik, kestirim sürecinin bir parçası olarak kayıp veri mekanizmasının modellenmesine gerek olmadığı anlamına gelmektedir.

Kayıp verilerin analiz dışı bırakılabilmesi için öncelikle kayıp verilerin ihmal edilebilir olduğunun kanıtlanması gerekmektedir. Bu tür bir kanıt, kayıp değerlerle, seçkisiz olarak belirlenen gözlenen verilerin örtüşeceği varsayımını destekler. Dolayısıyla tam ya da eksiksiz veri seti ile kayıp veri içeren veri setlerinin dağılımları ve bu veriler üzerinden yapılacak çıkarımlar arasında manidar bir fark olmadığı varsayımına ulaşılır (Rubin, 1976). Kayıp verilerin ihmal edilebilir olup olmaması, kayıp verilerin karakteristiği ve kayıp verilerin oluşmasına yol açan süreçlerle ilişkilidir. Bu noktada kayıp veri örüntüsü ve mekanizması, uygun analiz yöntemlerinin belirlenmesinde ve sonuçların yorumlanmasında anahtar rol oynamaktadır (Little & Rubin, 1987).

Kayıp veri mekanizmasının ihmal edilebilir olması, kayıp verilerin analiz dışı bırakılabilmesi ile araştırmacının işini oldukça kolaylaştıracaktır. Fakat aksi durumda kayıp veri mekanizmasının modellenmesi gerekir. İhmal edilemez kayıp veriler içeren veri seti, genellikle hangi modellerin daha uygun olacağı konusunda yeterli bilgi sağlayamaz. Araştırma sonuçlarının seçilen modele oldukça duyarlı olacağı da açıktır. Dolayısıyla ihmal edilebilir olmayan kayıp verilerle etkili bir kestirim yapılabilmesi için, kayıp veri sürecinin doğasına yönelik oldukça iyi ön bilgilere ihtiyaç vardır (Schaffer, 1997; Allison, 2002).

Kayıp verilerin varlığında, kullanılacak olan birçok istatistiksel yöntem geliştirilmiştir. Kullanılacak iyi bir yöntemin, 1) kayıp verilerden kaynaklanabilecek yanlılığı minimize etmesi, 2) elverişli bilgilerin kullanımını maksimize etmesi ve 3) belirsizliklere yönelik standart hata, güven aralığı ve olasılık değerlerine yönelik titiz kestirimler sağlaması beklenir (Allison, 2009).

Kayıp verilerle başa çıkmada kullanılan yöntemleri genel olarak iki grupta sınıflandırmak mümkündür: 1) Silmeye ve basit atamaya dayalı yöntemler, 2) olasılıklı ve ötelemeli veri atama yöntemleri. Olasılıklı ve ötelemeli yöntemler de kendi içerisinde iki grupta ele alınmaktadır: 1) En çok olabilirlik yaklaşımına dayalı yöntemler ve 2) çoklu atama yaklaşımına dayalı yöntemler. Bu sınıflama aynı zamanda kayıp veri yönteminin tarihsel gelişim aşamalarını da temsil etmektedir.

Kayıp verilere yönelik olarak ilk akla gelen ve alışıldık yöntemler, kayıp verilerin analiz dışı bırakılması ya da kayıp veriler yerine basit veri atama yöntemlerini içermektedir. Silmeye dayalı yöntemler içerisinde en bilindik olanlar 'dizin silme-DS (listwise deletion)', ve 'kısmî silme-KS (pairwise deletion)' yöntemleridir. Basit atamaya dayalı yöntemler içerisinde en bilindik olanı ise 'ortalama atama-OA (mean substitution)' yöntemidir. Ortalama atama, 'seri ortalaması atanması-SO (serial mean imputation)' şeklinde dikey olarak ya da 'gözlem birimleri ortalaması atanması-BO (unit's mean imputation)' şeklinde yatay olarak yapılabilmektedir. Ayrıca kayıp veriler yerine değer atamada, 'yakın noktalar ortalama atanması-YNO (mean of nearby points)' ve 'yakın noktalar medyan atanması-YNM (median of nearby points)' gibi bazı değerlerin ortalamalarını dikkate alan yöntemler de bulunmaktadır. Ancak bu yöntemin varyans-kovaryans kestiriminde yanlışlık ürettiği, artık iyi bilinen bir durumdur (Little & Rubin, 1987; Allison, 2002).

Silmeye ve basit değer atamaya dayalı bu yöntemler, istatistik yazılımlarında da yaygın bir şekilde kullanılmaktadır. Bununla birlikte bu yöntemlerin özellikle kayıp verilerin sınırlı miktarda olduğu durumlar dışında kullanılmaması önerilmektedir. Basit atama yöntemleri, veri atama sürecinde kayıp verilere yönelik belirsizlik olması durumunda, herhangi bir düzeltme sağlayamamaktadır. Bu nedenle basit veri atama yöntemlerinin 'dürüst olmayan (non honest)' yöntemler olduğu belirtilmektedir. Bu nedenle silme ve basit atamaya dayalı kayıp veri yöntemlerinin genellikle kayıp verilerin tam seçkisizlik özelliğini sağlaması durumunda kullanılabilir olduğu vurgulanmaktadır (Little & Rubin, 1987; Allison, 2002).

Silmeye ve basit değer atamaya dayalı geleneksel yöntemlerin eksiklikleri ve dezavantajları, kayıp verilere yönelik yeni yaklaşımlar üzerinde çalışılması ihtiyacını da beraberinde getirmiştir. Alternatif olarak geliştirilen pek çok yöntem arasında, 1990 başlarında teorik alt yapısı şekillendirilen ve 1990 sonlarında uygulama boyutuyla olgunlaştırılan 'en çok olabilirlik-EÇO (maximum likelihood)' ve 'çoklu veri atama-ÇVA (multiple imputation)' yaklaşımlarının öne çıktığı ve giderek daha yaygın bir şekilde kullanıldığı ifade edilmektedir (Dempster, Laird & Rubin, 1977; Rubin, 1987; Little & Rubin, 1987; Schaffer, 1997; Allison, 2002, 2009; Enders, 2010).

'En çok olabilirlik-EÇO (maximum likelihood)', bir yöntem olmaktan çok parametre kestiriminde olasılık temelli bir yaklaşımdır. Kayıp veri yöntemi olarak, bu yaklaşım temelinde geliştirilmiş 'doğrudan maksimum olabilirlik-DEÇO (direct maximum likelihood)' ya da 'beklenti-maksimizasyon algoritması-BM (expectation-maximization algorithm)' gibi bazı yöntemler bulunmaktadır (Rubin, 1987; Schaffer, 1997; Allison, 2002, 2009).

'Çoklu veri atama-ÇVA (multiple imputation)', her bir kayıp ya da eksik veri yerine, olasılıkların bir dağılımını yansıtan, kabul edilebilir iki ya da daha fazla verinin atanmasını öngören bir yaklaşımdır. ÇVA yaklaşımı temelinde geliştirilmiş 'tanımlayıcı atama-TA (deterministic imputation)', 'seçkisiz atama-SA (random imputation)' yöntemleri ya da bu ve benzeri yöntemleri kullanan 'Markov Chain Monte Carlo (MCMC)' yöntemi gibi daha özel yöntemler bulunmaktadır (Rubin, 1987; Schaffer, 1997; Allison, 2002, 2009).

Silme ve basit değer atamaya dayalı yöntemlere göre ÇVA ve EÇO yöntemlerinin en önemli üstünlüğü, bu yöntemlerin kayıp verilerin tam seçkisiz olarak oluşmadığı, kısmî seçkisizliğin sağlandığı durumlarda da kullanılabilmesidir. EÇO ve ÇVA yöntemleri, olasılıklı bir model üzerinden hareketle analizler öngörmektedir. Bu yöntemler için en kullanışlı model, çok değişkenli normal modeldir. Bununla birlikte bazı değişkenler normal dağılım varsayımını karşılamasa da çok değişkenli bir modelde oldukça başarılı kestirimler elde edilebildiği de belirtilmektedir (Rubin, 1987; Allison, 2002).

Kayıp verilerle başa çıkmaya yönelik yukarıda kısaca açıklanan genel yaklaşım ve yöntemlerin yanı sıra çoktan seçmeli testlerde yaygın olarak kullanılan iki yaklaşım daha görülmektedir. İki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde, kayıp verilerle başa çıkmada en yaygın yaklaşım, bu kayıp verilerin 'yanlış' olarak kabul edilmesidir. '0 Atama' yöntemi olarak tanımlanabilen bu yöntemin yaygın bir alternatifi ise kayıp verilerin 'uygulanmamış (not administered)' olarak kodlanması ve ileri analizlerde ihmal edilmesidir. Söz konusu bu her iki yaklaşımın da yanlı ve hatalı kestirimlere yol açtığı bilinmektedir. Bu yaklaşım ve yöntemler kayıp verilere yönelik optimal birer çözüm olarak görülmemektedir. Ancak boş bırakılan maddelerin tamamen seçkisiz oluşması koşulunun sağlanması durumunda, yanıtlama yapan ve yapmayan bireylere yönelik kestirimlerin karşılaştırılması mümkün olabilmektedir. Bu nedenle bu tür kayıp verilere yönelik olarak geliştirilmiş daha karmaşık kayıp veri yöntemlerinin kullanımı gerekli görülmektedir (Lord, 1974; Hohensinn & Kubinger, 2011; Culbertson, 2011).

Özellikle iki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde, yanıtlayıcının bilinçli bir şekilde boş bıraktığı ya da atladığı maddelere bağlı olarak oluşan kayıp verilerin ihmal edilebilir olmadığı, üzerinde uzlaşılan bir belirlemedir. Bu tür kayıp verilerin varlığında, elde edilen kestirimlerin manidar düzeyde yanlılık içerme olasılığı yüksektir. Yanlılık olasılığını düşürmede en yüksek performans gösteren kayıp veri yöntemlerinin ise EÇO ve ÇVA yöntemleri olduğu sıklıkla belirtilmektedir (Elashoff & Elashoff, 1971; Lord, 1983; Mislavy & Wu, 1996; Schaffer, 1997; DeMars, 2002; Linacre, 2004; Finch, 2008; Hohensinn & Kubinger, 2011).

Kayıp veri sorununa yönelik olarak geliştirilmiş pek çok çözüm yönteminden bazılarının, diğerlerine göre daha 'iyi' olduğu ileri sürülebilir. Fakat bu yöntemlerden hiç birisi gerçek anlamda 'iyi' olarak tanımlanamaz. Gerçek çözüm, kayıp veri olmaması ya da kayıp veri miktarının ihmal edilebilir düzeyde olmasıdır. Bu nedenle bir araştırmada kayıp veri miktarının en aza indirgenmesi çabası önemlidir. Özensiz ve dikkatsiz bir şekilde yürütülen bir araştırmada herhangi bir istatistiksel düzeltmenin bir anlamı da yoktur (Allison, 2002).

Araştırmalarda kayıp veri sorununun ancak 1970 sonrasında tartışılmaya başlanması ve kayıp veri yöntemlerinin ancak 1990 sonrasında istatistiksel yazılımlara yansması, geliştirilmiş birçok kayıp veri yönteminin henüz beklentileri karşılayacak düzeyde olmamasını bir ölçüde açıklamaktadır. Normallik varsayımının sağlanmadığı modellerde, ihmal edilebilir olmayan kayıp verilerin varlığında, kategorik ve iki kategorili puanlanan değişkenlerde, çok değişkenli ve çok düzeyli analizlerin büyük çoğunluğunda kayıp veriler, hâlâ önemli bir sorun olarak görülmektedir.

Çok kategorili puanlanan maddelerden oluşan testlerde ya da sürekli değişkenlerde kayıp verilerle başa çıkabilme, üzerinde daha fazla çalışılmış bir alandır ve bu tür verilere yönelik olarak görece daha fazla yöntem geliştirildiği görülmektedir. İki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde ise kayıp verilerle başa çıkabilme, madde ve test karakteristiklerinin doğal bir sonucu olarak daha zordur. Bu tür veri

setlerinde kullanılabilir kayıp veri yöntemleri ve bu yöntemlerin, olası yanlışlık riskini minimize etmedeki performans ve yeterlilikleri, halen üzerinde tartışılan bir konu olarak güncelliğini korumaktadır.

Bu çalışmanın amacı; kayıp verilerin varlığında, çoktan seçmeli testlerde, farklı kayıp veri yöntemleri kullanılarak kestirilen madde ve test parametreleri arasındaki ilişkilerin incelenmesi ve bu tür testlerde kullanılması uygun olan kayıp veri yöntemlerinin belirlenmesidir.

Bu çalışma, iki kategorili puanlanan verilerde kullanılabilir uygun olan ve uygun olmayan kayıp veri yöntemlerine yönelik kanıt sağlaması açısından önemli görülmektedir. Ayrıca analizlerin simülatif veriler yerine gerçek bir veri seti üzerinde gerçekleştirilmesi, kayıp veri yöntemlerinin performansının değerlendirilmesinde bir avantaj olarak değerlendirilmektedir. Bu çalışmanın, ilgili alanyazına katkı sağlamanın yanı sıra özellikle iki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde kayıp veri sorununa dikkat çekmesi beklenmektedir.

YÖNTEM

Araştırma Modeli

Bu çalışma, temel araştırma türünde, ilişkisel tarama modelinde bir araştırma olarak tasarlanmıştır. Farklı kayıp veri yöntemleri, bu yöntemler kullanılarak elde edilen madde ve test parametreleri arasındaki ilişkiler üzerinden incelenmiştir.

Evren

Bu çalışmanın evreni, SBS 2011 Matematik Testi A Kitapçığını alan 527517 yanıtlayıcıdan oluşmaktadır. Kayıp verilere yönelik inceleme ve analizler, evrenden elde edilen veri seti üzerinde yürütülmüştür.

Veri Toplama Aracı

Bu çalışmada veri toplama aracı olarak, SBS 2011 Matematik Testi A Kitapçığı kullanılmıştır. SBS Matematik Testi, matematik başarısına yönelik çoktan seçmeli ve dört seçenekli 20 maddeden oluşmaktadır. Test sonuçlarının elde edilmesinde yanıtlayıcı tepkileri, 1-0 şeklinde yapay süresiz olarak puanlanmaktadır.

SBS, Türkiye genelinde, Millî Eğitim Bakanlığı (MEB) tarafından, ilköğretim kademesinin 8. sınıfından mezun olma aşamasında olan öğrencilere yönelik olarak gerçekleştirilen ve her yıl düzenlenen bir başarı testidir. Test sonucunda, yanıtlayıcıların başarı puanları, okul başarıları da dikkate alınarak bir üst eğitim kademesine geçişte, seçme ve yerleştirme amacıyla kullanılmaktadır.

SBS, Türkçe Testi, Matematik Testi, Fen Bilimleri Testi, Sosyal Bilimler Testi ve Yabancı Dil Testi olmak üzere beş alt testten oluşmaktadır. Bu testlere yönelik maddelerin yazımında, ilgili ders programında yer alan konu ve kazanımların dikkate alındığı belirtilmektedir. Her ne kadar test geliştirme süreçlerine uygunluğu açısından eleştirilebilecek olsa da SBS testlerinin geliştirilmesi ve uygulanması süreçlerinin kendi içerisinde bir mantığı olduğu anlaşılmaktadır³.

³ SBS ve Ortaöğretime Geçiş Sistemi hakkında ayrıntılı bilgi için Millî Eğitim Bakanlığının ilgili internet sitesine; http://oges.meb.gov.tr/sbs_istat.htm adresinden erişilebilmektedir.

Verilerin Analizi

SBS 2011 Matematik Testi A Kitapçığına yönelik ham veri seti, testi alan 527517 yanıtlayıcıya yönelik A-B-C-D şeklinde kodlanan yanıt örüntüleri ile cinsiyet, okul türü ve il değişkenlerini içermektedir. Cinsiyet değişkeni, 'erkek' ve 'kız' olmak üzere iki kategorilidir. Okul türü değişkeni, 'devlet' ve 'özel okul' olmak üzere iki kategorilidir. İl değişkeni ise Türkiye'nin 81 iline göre kategorik bir değişken olarak tanımlıdır.

Ham veri setinde, bu çalışmada öngörülen analizler dikkate alınarak gerekli ön düzenlemeler yapılmıştır. Öncelikle testin cevap anahtarı dikkate alınarak, yanıt örüntüleri 1-0 örüntüsüne dönüştürülmüştür. Bu dönüştürmede doğru yanıtlar 1, yanlış yanıtlar 0 olarak puanlanmış, kayıp veriler ise özel kodlama ile tanımlanmıştır. İkinci olarak ham veri setinde yer alan il değişkeni, Türkiye 'İstatistikî Bölge Birimleri Sınıflaması (İBBS) Düzey 1' dikkate alınarak, 12 bölge esasına göre yeniden kodlanmış ve 'bölge' değişkeni olarak veri setine yerleştirilmiştir. Üçüncü olarak 'toplam doğru yanıt sayıları', 'toplam yanlış yanıt sayıları' ve 'kayıp veri içeren madde sayıları' her bir gözlem birimi düzeyinde hesaplanmış, değişken olarak veri setine yerleştirilmiştir. Son olarak gözlem birimleri düzeyinde kayıp veri bulunup bulunmaması durumu, bir 'kukla (dummy) değişken' olarak tanımlanmıştır. Böylelikle ham veri seti, 1-0 yanıt örüntülerinin yanı sıra 7 bağımsız değişken içerecek şekilde ileri analizlere uygun hale getirilmiştir.

Ham veri setinin ileri analizlere uygun hale getirilmesinden sonra öngörülen ileri analizler için temel varsayımların sağlanıp sağlanmadığı incelenmiş ve test edilmiştir. Kayıp veri yöntemlerinden herhangi birinin kullanılıp kullanılmayacağı ya da uygun kayıp veri yönteminin hangisi olduğu kararı, kayıp veri örüntüsü ve kayıp veri mekanizmasına yönelik inceleme ve analizlere bağlı olarak verilmektedir. Bu nedenle öncelikle kayıp veri örüntüsü ve kayıp veri mekanizmasına yönelik inceleme ve analizler gerçekleştirilmiştir.

Kayıp veri örüntüsüne yönelik inceleme ve analizlerle kayıp verilerin veri setindeki konumlanma biçiminin betimlenmesi ve tanımlanması amaçlanmıştır. Doğru ve yanlış yanıt sayıları ile kayıp veri içeren madde sayılarının dağılımlarına yönelik betimsel istatistikler, ayrıca yanıt örüntülerinin sıklığı dikkate alınarak yürütülen inceleme ve analizler sonucunda kayıp verilerin gerek maddeler düzeyinde gerekse yanıt örüntüleri düzeyinde düzenli, planlı ya da monoton bir dağılım göstermediği, kayıp veri örüntüsünün, 'genel örüntü' yapısına uygun olduğu belirlenmiştir.

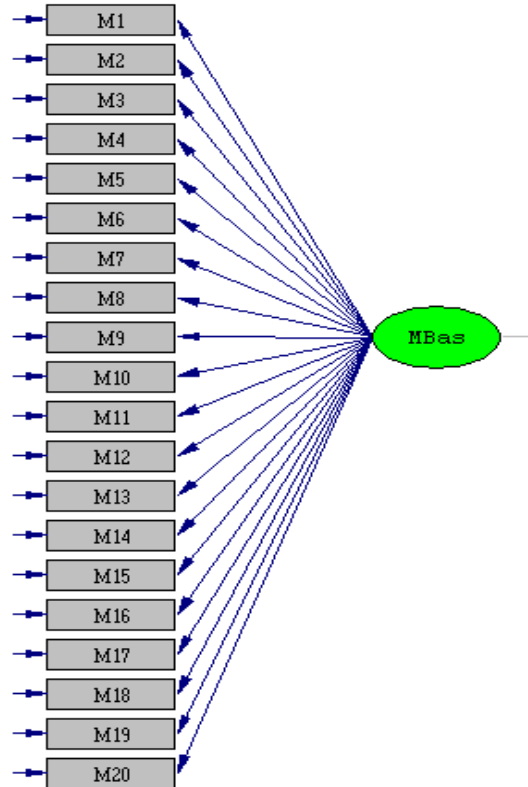
Kayıp veri mekanizmasına yönelik inceleme ve analizlerle, kayıp verilerin oluşmasına yol açan mekanizmanın tam seçkisizlik ya da kısmî seçkisizlik özelliklerine sahip olup olmadığının belirlenmesi amaçlanmıştır. Bu mekanizmalar, özellikle iyi tanımlanmış değişkenler olmaksızın, doğrudan test edilebilir olmamakla birlikte bir takım istatistiksel kanıtlardan hareketle değerlendirilebilir görülmektedir. Kayıp veri mekanizmasının tanımlanmasına yönelik istatistiksel kanıtlar genellikle kayıp veri mekanizması ile ilişkili olmaması beklenen değişkenlerin kategorilerine göre yanıt örüntülerinde manidar bir fark oluşup oluşmadığının incelenmesi yaklaşım ve yöntemlerini içermektedir. Sıklıkla başvurulan diğer bir yöntem ise genel yanıtlayıcı kitlesi ile yanıt örüntülerinde kayıp veri bulunmayan yanıtlayıcı kitlesinin profillerinin karşılaştırılmasıdır.

Bu çalışmada, kayıp veri mekanizması, cinsiyet, okul türü ve bölge değişkenleri dikkate alınarak gerçekleştirilen analizlerle tanımlanmaya çalışılmıştır. Cinsiyet, okul türü ve bölge değişkenlerinin kategorileri arasında, yanıtlayıcıların, doğru ve yanlış yanıt sayıları ile kayıp veri içeren madde sayılarının manidar düzeyde farklılaşp farklılaşmadığı test edilmiştir. Bu incelemede, yanıtlayıcıların tamamının yanı sıra yanıt örüntüleri gözlenen verilerden oluşan, madde düzeyinde kayıp veri içeren ve birim düzeyinde kayıp veri içeren

yanıtlayıcılar da ayrı ayrı gruplar olarak dikkate alınmıştır. Kayıp veri mekanizmasına yönelik inceleme ve analizler, kayıp verilerin oluşmasını açıklayan mekanizmanın tam seçkisizlik özelliğini sağlamadığını fakat kısmi seçkisizlik özelliğini sağlama olasılığının yüksek olduğunu göstermiştir.

Kayıp veri örüntüsü ve kayıp veri mekanizmasına yönelik inceleme ve analizlerin yanı sıra istatistiksel yazılımların özellikleri ve erişilebilirlikleri de dikkate alınarak kullanılabilir 12 kayıp veri yöntemi belirlenmiştir. Bu yöntemlerden 'dizin silme (DS)' yöntemi silmeye dayalı, '0 Atama', 'seri ortalamaları ataması (SO)', 'gözlem birimi ortalama ataması (BO)', 'yakın noktalar ortalama ataması (YNO)', 'yakın noktalar medyan ataması (YNM)', 'doğrusal interpolasyon (Dİ)' ve 'doğrusal eğilim noktası ataması (DE)' yöntemleri basit atamaya dayalı, 'regresyon atama (RA)', 'beklenti-maksimizasyon algoritması (BM)' ve 'veri çoğaltma (VÇ)' yöntemleri en çok olasılık yaklaşımına dayalı, 'Markov zincirleri Monte Carlo (MCMC)' ise çoklu atama yaklaşımına dayalı yöntemlerdir.

İleri analizlerde kullanılacak kayıp veri yöntemlerinin belirlenmesinden sonra belirlenen her bir kayıp veri yönteminde de testin yapı geçerliği açısından yeterli düzeyde olup olmadığı ve özellikle tek boyutluluğun sağlanıp sağlanmadığı, 'doğrulayıcı faktör analizi (DFA)' ile test edilmiştir. DFA, sürekli verilerde, kovaryans ya da korelasyon matrisleri üzerinden hareketle kestirimler üretmektedir. İki kategorili puanlanan veri setlerinde ise DFA, asimetrik kovaryans matrisleri üzerinden hesaplama yapmaktadır. Bu nedenle bu çalışmada DFA, asimetrik kovaryans matrisleri tanımlanarak uygulanmıştır. Matematik başarısının 'gizil değişken', her bir maddenin 'gözlenen değişken' olarak tanımlanması ile üretilen ölçme modeli Grafik 1'de, kayıp veri yöntemlerine göre kestirilen uyum iyiliği indeksleri ise Tablo 1'de gösterilmektedir.



Grafik 1. SBS 2011 matematik testi A kitapçığı verilerine yönelik ölçme modeli

Tablo 1. Kayıp veri yöntemlerine göre DFA model uyum iyiliği parametreleri

Kayıp Veri Yöntemi	Hata Varyansları		χ^2/sd^{***}	NFI	GFI	AGFI	RMSEA	RMR	SRMR
	Yol Katsayıları Aralığı**	Yol Katsayıları Aralığı**							
DS	.37 - .76	.42 - .86	2.09	.99	.98	.97	.028	.0051	.021
0 Atama	.30 - .64	.59 - .91	8.36	.98	.97	.97	.038	.0056	.028
BO	.32 - .71	.50 - .90	2.02	.99	.98	.97	.026	.0050	.022
SO	.27 - .63	.60 - .93	6.15	.98	.98	.98	.032	.0042	.024
YNO	.29 - .64	.59 - .92	7.89	.98	.97	.97	.037	.0055	.027
YNM	.30 - .65	.58 - .96	11.76	.97	.96	.95	.046	.0073	.036
Dİ	.27 - .62	.62 - .93	5.58	.98	.98	.98	.030	.0046	.024
DE	.27 - .63	.60 - .93	6.13	.98	.98	.98	.032	.0042	.024
RA	.26 - .65	.57 - .93	8.41	.98	.97	.97	.038	.0058	.028
BM	.28 - .69	.53 - .92	11.34	.98	.96	.96	.045	.0056	.031
VÇ*	.26 - .63	.60 - .93	20.30	1.00	.99	.99	.019	.0031	.013
MCMC*	.26 - .66	.57 - .93	21.81	.99	.99	.98	.029	.0044	.020

* Model için önerilen modifikasyonlar dikkate alınarak hata varyansları ilişkilendirilmiştir.

** Standartlaştırılmış değerler dikkate alınmıştır ($t > 2.56$, $p < .01$).

*** $p = .000$ hesaplanmıştır.

Tablo 1'de verilen DFA kestirimlerine göre standartlaştırılmış yol katsayıları genel olarak .30'un üzerindedir. Ayrıca yol katsayılarının manidarlığına yönelik t değerlerinin tamamı 2.56'nın üzerindedir. Bu durum gözlenen ve gizil değişkenler arasında tanımlı tüm yolların istatistiksel olarak .01 düzeyinde manidar olduğunu göstermektedir. Hata varyansları genel olarak .90'ın altındadır. Örneklem büyüklüğüne fazlaca duyarlı olan χ^2/sd oranları, DS ve BO yöntemlerinin kullanılması durumunda mükemmel düzeyde bir model-veri uyumunun elde edildiğini, YNM, BM, VÇ ve MCMC yöntemlerinde model-veri uyumu bulunmadığını, diğer yöntemlerin ise zayıf düzeyde model-veri uyumu sağlayabildiğini göstermektedir. Diğer taraftan uyum iyiliği parametrelerinden NFI, GFI ve AGFI değerleri, genel olarak .97 ve üzerindedir. Hata parametrelerinden RMSEA, RMR ve SRMR değerleri ise genel olarak .05'in altında ve 0'a oldukça yakındır. Bu değerler her bir kayıp veri yöntemine göre mükemmel düzeyde model-veri uyumunun sağlandığını göstermektedir.

SBS 2011 Matematik Testi A Kitapçığına yönelik veri seti üzerinde yapı geçerliği çalışmaları olarak yürütülen DFA sonuçları, her bir kayıp veri yönteminde de maddelerin tek bir faktör altında toplanabildiğini, tek faktörlü yapının mükemmel yakın düzeyde model-veri uyumu sağladığını göstermiştir. Bu durum testin tek boyutluluk özelliğine sahip olduğuna ve ileri analizlerde toplam puanların kullanılabilmesine dair yeterli düzeyde kanıt oluşturmaktadır. Bununla birlikte genel olarak silme ve basit atamaya dayalı yöntemlerin, en çok olabilirlik ve çoklu atamaya dayalı yöntemlere göre DFA'da daha iyi performans gösterdiği ileri sürülebilir.

İleri analizlere yönelik temel varsayımların test edilmesinden sonra belirlenen 12 farklı kayıp veri yöntemine göre ayrı ayrı eksiksiz veri setleri elde edilmiştir. Elde edilen eksiksiz veri setleri üzerinde, 'klasik test kuramı (KTK)' temelinde madde ve test parametre kestirimleri gerçekleştirilmiştir. Madde parametreleri olarak madde güçlük indeksi, madde ayırıcılık indeksi ve madde güvenilirlik katsayısı, test parametreleri olarak test ortalama güçlüğü, test ortalama güçlük indeksi, test ortalama ayırıcılık indeksi ve test güvenilirlik katsayısı kestirimleri gerçekleştirilmiştir.

Farklı kayıp veri yöntemleri kullanılarak elde edilen eksiksiz veri setleri üzerinde kestirilen madde ve test parametreleri arasındaki ilişkiler 'Spearman Sıra Farkları Korelasyon Katsayısı' ile incelenmiştir. Kestirilen parametreler arasındaki farkların manidarlığı ise parametrik olmayan testlerden 'Friedman'ın sıra farklarına dayalı ilişkili örneklem çift yönlü varyans analizi' ve 'Kendall'in ilişkili örneklemler konkordans katsayısı' kestirimleri ile test edilmiştir.

Bu çalışmada yürütülen analizlerde özel amaçlı istatistiksel yazılımlardan yararlanılmıştır. Kullanılan kayıp veri yöntemlerinin tamamını içeren bir istatistiksel yazılım bulunmamaktadır. Bu nedenle, verilerin analizinde, belirlenen kayıp veri yöntemlerinin kullanımına uygun olarak, SPSS 15.0, SPSS 20.0, LISREL 8.7, NORM ve IMPUTE yazılımları kullanılmıştır. Madde ve test parametrelerinin kestirimleri ise ITEMANW yazılımı ile gerçekleştirilmiştir.

BULGULAR

SBS 2011 Matematik Testi A Kitapçığını alan yanıtlayıcılara yönelik gözlenen ve kayıp verilerin dağılımı Tablo 2'de verilmektedir.

Tablo 2. Gözlenen ve kayıp verilerin dağılımı

	N	%
Gözlenen Veri	144553	27.40
Kayıp Veri (Madde Düzeyinde)	382937	72.59
Kayıp Veri (Birim Düzeyinde)	27	.01
Toplam	527517	100.00

Tablo 2'de görüldüğü gibi testi alan gözlem birimi yani yanıtlayıcı sayısı 527517'dir. Yanıtlayıcıların % 27.40'ı matematik testinde yer alan maddelerin her birinde, doğru ya da yanlış, yanıtlama yapmıştır. Bu yanıtlayıcılara yönelik yanıt örüntülerinde kayıp veri bulunmamaktadır. Bu durumda kayıp veri içermeyen yani eksiksiz alt veri seti 144553 gözlem biriminden oluşmaktadır. Yanıtlayıcıların 382964'ü (% 72.60) testte yer alan maddelerden en az birinde yanıtlama yapmamıştır. Bu yanıtlayıcıların neredeyse tamamının (% 79.59) yanıt örüntülerinde madde düzeyinde kayıplar bulunmaktadır. Birim düzeyindeki kayıplar ise tüm gözlem birimlerinin % 0.01'ini oluşturmaktadır. Bu yanıtlayıcılar, testte yer alan maddelerin tamamında yanıtlama yapmamıştır ya da birden fazla seçenek işaretleme, belirgin olmayan şekilde işaretleme gibi gerekçelerden dolayı yanıtları geçersiz sayılmıştır.

Rubin (1987), Puma ve diğerleri (2009), birim düzeyinde kayıp verilerin örnekleme hatasına, madde düzeyinde kayıp verilerin ise madde yanlışına işaret edebileceğini belirtmektedir. SBS 2011 Matematik Testi A Kitapçığına yönelik verilerde birim düzeyinde kayıplar, oldukça azdır. Bu durumda, SBS'nin Türkiye genelinde uygulanan bir test olduğu da dikkate alındığında, kayıp verilerin oluşmasının örnekleme hatasından kaynaklanma olasılığı düşüktür. Diğer taraftan madde düzeyinde kayıp veri miktarının fazlalığı, madde yanlışlığı olasılığının yüksek olduğunu göstermektedir.

SBS 2011 Matematik Testi A Kitapçığında yer alan her bir maddede doğru yanıtlama, yanlış yanıtlama sayıları ve kayıp veri sayılarının dağılımları Tablo 3'te verilmektedir.

Tablo 3. Maddeler düzeyinde tepkilerin dağılımı

Madde	Doğru		Yanlış		Kayıp		Toplam	
	N	%	N	%	N	%	N	%
1	131528	24.9	319443	60.6	76546	14.5	527517	100.0
2	131145	24.9	313670	59.5	82702	15.7	527517	100.0
3	147484	28.0	215742	40.9	164291	31.1	527517	100.0
4	160423	30.4	185470	35.2	181624	34.4	527517	100.0
5	228436	43.3	174604	33.1	124477	23.6	527517	100.0
6	130361	24.7	355921	67.5	41235	7.8	527517	100.0
7	213151	40.4	194275	36.8	120091	22.8	527517	100.0
8	235407	44.6	210916	40.0	81194	15.4	527517	100.0
9	155951	29.6	311763	59.1	59803	11.3	527517	100.0
10	184155	34.9	255874	48.5	87488	16.6	527517	100.0
11	190659	36.1	208363	39.5	128495	24.4	527517	100.0
12	167622	31.8	271045	51.4	88850	16.8	527517	100.0
13	84776	16.1	377078	71.5	65663	12.4	527517	100.0
14	156593	29.7	280101	53.1	90823	17.2	527517	100.0
15	211270	40.0	268379	50.9	47868	9.1	527517	100.0
16	173559	32.9	303382	57.5	50576	9.6	527517	100.0
17	323179	61.3	158305	30.0	46033	8.7	527517	100.0
18	138864	26.3	241590	45.8	147063	27.9	527517	100.0
19	89559	17.0	313571	59.4	124387	23.6	527517	100.0
20	198427	37.6	318652	60.4	10438	2.0	527517	100.0

Tablo 3'te görüldüğü gibi gözlenen veriler dikkate alındığında, test maddeleri düzeyinde doğru yanıtlama yüzdeleri % 16.1 ile % 61.3 arasında değişmektedir. En fazla doğru yanıtlanan madde 17. maddedir. Bunu sırasıyla 8, 5, 7 ve 15. maddeler izlemektedir. En az doğru yanıtlanan madde ise 13. maddedir. Bunu sırasıyla 19, 6, 1 ve 2. maddeler izlemektedir. Doğru yanıtlama yüzdelerinde düzenli bir artış ya da azalış görülmemektedir.

Test maddeleri düzeyinde yanlış yanıtlama yüzdeleri % 30.0 ile % 71.5 arasında değişmektedir. En az yanlış yanıtlanan madde 17. maddedir. Bunu sırasıyla 5, 4, 7 ve 11. maddeler izlemektedir. En fazla yanlış yanıtlanan madde 13. maddedir. Bunu sırasıyla 6, 1, 20 ve 2. maddeler izlemektedir. Yanlış yanıtlama yüzdelerinde, düzenli bir artış ya da azalış görülmemektedir.

Testte yer alan 20 maddenin her biri, kayıp veri içermektedir. Bu maddelerden 19'unda % 5'in üzerinde, 14'ünde % 10'un üzerinde, 7'sinde ise % 20'nin üzerinde kayıp veri bulunmaktadır. Test maddeleri düzeyinde kayıp veri yüzdeleri % 2.0 ile % 34.4 arasında değişmektedir. En az kayıp veri içeren madde 20. maddedir. Bunu sırasıyla 6, 17, 15 ve 16. maddeler izlemektedir. En fazla kayıp veri içeren madde ise 4. maddedir. Bunu sırasıyla 3, 18, 11, 5 ve 19. maddeler izlemektedir. Kayıp veri yüzdelerinde, düzenli bir artış ya da azalış görülmemektedir.

Madde Parametreleri

SBS 2011 Matematik Testi A Kitapçığına yönelik madde parametrelerinden madde güçlük indeksi, madde ayırıcılık indeksi ve madde güvenilirlik katsayıları, her bir kayıp veri yöntemine göre ayrı ayrı kestirilmiş ve aşağıda sunulmuştur.

SBS 2011 Matematik Testi A Kitapçığında yer alan maddelere yönelik olarak, her bir kayıp veri yöntemine göre ayrı ayrı elde edilen madde güçlük indeksi kestirimleri Tablo 4'te verilmektedir.

Tablo 4. Kayıp veri yöntemlerine göre madde güçlük indeksleri

Madde	Kayıp Veri Yöntemleri											
	DS	0 Atama	BO	SO	YNO	YNM	Dİ	DE	RA	BM	VÇ	MCMC
1	.38	.25	.28	.25	.25	.25	.32	.25	.29	.27	.29	.29
2	.40	.25	.28	.25	.25	.25	.31	.25	.28	.26	.28	.28
3	.46	.28	.35	.28	.29	.28	.44	.28	.39	.34	.39	.39
4	.53	.30	.38	.30	.32	.30	.49	.30	.45	.38	.44	.45
5	.54	.45	.47	.67	.48	.67	.61	.67	.55	.53	.55	.55
6	.31	.23	.25	.23	.23	.23	.26	.23	.25	.23	.26	.25
7	.52	.41	.45	.63	.43	.63	.56	.63	.52	.50	.52	.52
8	.55	.45	.47	.59	.46	.59	.55	.59	.52	.51	.52	.52
9	.41	.29	.34	.29	.30	.29	.35	.29	.32	.31	.32	.32
10	.46	.36	.41	.36	.37	.36	.45	.36	.42	.40	.42	.42
11	.49	.35	.40	.35	.37	.35	.50	.35	.44	.42	.44	.45
12	.44	.31	.34	.31	.32	.31	.40	.31	.36	.34	.36	.36
13	.32	.17	.20	.17	.17	.17	.21	.17	.20	.17	.19	.19
14	.43	.30	.32	.30	.30	.30	.38	.30	.35	.31	.35	.35
15	.47	.39	.39	.39	.40	.39	.45	.39	.42	.40	.42	.42
16	.43	.34	.35	.34	.34	.34	.39	.34	.37	.35	.37	.37
17	.59	.61	.62	.70	.63	.70	.69	.70	.66	.67	.66	.66
18	.42	.27	.35	.27	.27	.27	.41	.27	.36	.31	.37	.36
19	.35	.16	.24	.16	.17	.16	.25	.16	.22	.18	.22	.21
20	.46	.39	.40	.39	.39	.39	.40	.39	.39	.39	.39	.39

Tablo 4'te görüldüğü gibi farklı kayıp veri yöntemleriyle elde edilen madde güçlük indeksi kestirimleri arasında betimsel farklılıklar bulunmaktadır. Bu betimsel farklılıklara yönelik olarak farklı kayıp veri yöntemlerine göre kestirilen madde güçlük indeksleri arasında, pozitif yönlü ve mükemmel yakın düzeyde manidar ilişkiler bulunduğu belirlenmiştir (Sperman rho katsayıları .85-.99 arasında ve $p < .01$). Friedman'ın sıra farklarına dayalı ilişkili örneklem çift yönlü varyans analizi sonuçları ($p < .01$) ve Kendall'in ilişkili örneklemler konkordans katsayısı kestirimleri ($p < .01$), genel olarak her bir kayıp veri yöntemi grubuna göre elde edilen kestirimler arasında .01 düzeyinde manidar fark bulunduğunu göstermektedir. Buna göre 0 Atama yöntemi, silmeye dayalı yöntem, basit atamaya dayalı yöntemler ve bir grup olarak MVA-ÇVA yöntemleri ile elde edilen kestirimler arasındaki farklar manidardır.

Genel olarak en yüksek kestirimler DS yönteminde, en düşük kestirimler ise 0 Atama yönteminde elde edilmiştir. Diğer bir deyişle DS yönteminin kullanılması durumunda maddelerin 'kolay madde', 0 Atama yönteminin kullanılması durumunda ise maddelerin 'zor madde' olarak değerlendirilme olasılığı yüksektir.

Her bir kayıp veri yönteminde de madde güçlük indeksi en yüksek olan madde 17. maddedir. Bu maddeye yönelik güçlük indeksi değerleri, farklı kayıp veri yöntemlerine göre .61 ve .70 arasında değişmektedir. Madde güçlük indeksi en düşük olan madde ise DS yönteminde 6. madde, diğer yöntemlerde 13. ve 19. maddelerdir.

Bir teste yer alacak maddelerin güçlük düzeyinin ne olması gerektiği, testin amacına göre değerlendirilmektedir. Genellikle .20 ile .80 arasındaki güçlük değerleri, maddenin teste alınabilir olduğuna yönelik bir kanıt sağlamaktadır. Yanıtlayıcının maksimum performans göstermesi beklenen testlerde, güçlük indeksleri görece daha düşük yani 'zor' ya da 'çok zor' maddeler kullanılabilir. Eğitim alanında özellikle öğrenci başarısını belirlemeye

yönelik testlerde ise maddelerin çoğunluğunun güçlüklerinin orta düzeyde ve .50 civarında olması önerilmektedir (Lord & Novick, 1968; Özçelik, 1981; Crocker & Algina, 1986; Baykul, 2000; Tekin, 2007).

Madde güçlük indekslerinin .20 ile .80 aralığında olması ölçütü dikkate alındığında bu beklentiyi karşılayan yöntemlerin DS, BO, Dİ ve RA yöntemleri olduğu görülmektedir. Hiçbir kayıp veri yönteminde, madde güçlük indeksi .80'in üzerinde olan madde bulunmamaktadır. Diğer taraftan indeks değeri .20'nin altında olan madde sayısı VÇ ve MCMC yöntemlerinde 1, 0 Atama, SO, YNO, YNM, DE ve BM yöntemlerinde ise 2'dir.

Madde güçlük indekslerinin .50 civarında olması ölçütü dikkate alındığında, bu beklentiyi en iyi karşılayan yöntemlerin DS ve Dİ yöntemleri olduğu görülmektedir. Madde güçlük indeksleri .16 ile .70 arasında değerler almaktadır. İndeks değerleri aralığı en düşük olan yöntem DS yöntemi (.31 - .59 arası), en yüksek olan yöntemler ise SO, YNM ve DE yöntemleridir (.19 - .70).

SBS 2011 Matematik Testi A Kitapçığında yer alan maddelere yönelik olarak, her bir kayıp veri yöntemine göre ayrı ayrı elde edilen madde ayırıcılık indeksi kestirimleri Tablo 5'te verilmektedir.

Tablo 5. Kayıp veri yöntemlerine göre madde ayırıcılık indeksleri

Madde	Kayıp Veri Yöntemleri											
	DS	0 Atama	BO	SO	YNO	YNM	Dİ	DE	RA	BM	VÇ	MCMC
1	.80	.55	.68	.50	.52	.50	.60	.50	.58	.57	.59	.60
2	.87	.57	.71	.52	.54	.52	.59	.52	.60	.59	.61	.61
3	.79	.53	.81	.48	.50	.48	.60	.48	.63	.68	.62	.64
4	.62	.36	.66	.30	.33	.30	.40	.30	.43	.52	.41	.42
5	.79	.72	.80	.45	.65	.45	.62	.45	.66	.72	.67	.68
6	.47	.34	.39	.31	.30	.31	.30	.31	.31	.29	.31	.31
7	.80	.70	.83	.53	.64	.53	.64	.53	.72	.78	.69	.73
8	.59	.43	.54	.33	.38	.33	.40	.33	.39	.46	.41	.40
9	.88	.65	.75	.60	.63	.60	.67	.60	.69	.67	.70	.69
10	.88	.74	.87	.70	.71	.70	.76	.70	.80	.83	.79	.81
11	.81	.68	.84	.61	.64	.61	.68	.61	.72	.78	.71	.73
12	.87	.63	.78	.57	.59	.57	.65	.57	.67	.68	.68	.68
13	.52	.24	.32	.21	.22	.21	.23	.21	.23	.21	.23	.23
14	.64	.48	.60	.43	.44	.43	.46	.43	.47	.47	.46	.47
15	.83	.70	.71	.66	.65	.66	.65	.66	.71	.69	.71	.70
16	.77	.61	.72	.58	.58	.58	.61	.58	.63	.63	.64	.63
17	.74	.63	.67	.51	.54	.51	.54	.51	.61	.55	.60	.61
18	.74	.41	.68	.37	.37	.37	.47	.37	.50	.50	.49	.50
19	.78	.31	.60	.27	.29	.27	.38	.27	.40	.35	.39	.39
20	.83	.60	.64	.59	.57	.59	.59	.59	.62	.59	.62	.62

Tablo 5'te görüldüğü gibi kullanılan kayıp veri yöntemine göre madde ayırıcılık indeksleri kestirimlerinde betimsel farklılıklar oluşmaktadır. Bu betimsel farklılıklara yönelik olarak farklı kayıp veri yöntemlerine göre kestirilen madde ayırıcılık indeksleri arasında, pozitif yönlü, çoğunlukla orta ve güçlü düzeyde manidar ilişkiler bulunduğu belirlenmiştir (Sperman rho katsayıları .70-.99 arasında ve $p < .01$). Friedman'ın sıra farklarına dayalı ilişkili örneklem çift yönlü varyans analizi sonuçları ($p < .01$) ve Kendall'in ilişkili örneklemler konkordans katsayısı kestirimleri ($p < .01$), genel olarak her bir kayıp veri yöntemi grubuna göre elde edilen kestirimler arasında .01 düzeyinde manidar fark bulunduğunu göstermektedir. Buna göre 0 Atama yöntemi, silmeye dayalı yöntem, basit atamaya dayalı

yöntemler ve bir grup olarak MVA-ÇVA yöntemleri ile elde edilen kestirimler arasındaki farklar manidardır.

Genel olarak en yüksek madde ayırıcılık indeksleri DS yönteminde, en düşük ayırıcılık indeksleri ise SO ve YNM yöntemlerinde elde edilmektedir. Madde ayırıcılık indeksleri, DS yönteminde .47 ile .88 arasında, SO ve YNM yöntemlerinde ise .21 ile .70 arasında değişmektedir. Ayırıcılık indeksleri aralığı en dar olan yöntem DS yöntemidir. Aralığın en geniş olduğu yöntem ise BM yöntemidir. BM yönteminde madde ayırıcılık indeksleri .21 ile .86 arasında değerler almaktadır.

Ayırıcılık indeksi en yüksek olan madde, DS yönteminde 9 ve 10. maddeler, diğer yöntemlerde 10. maddedir. İndeks değeri en düşük olan madde ise DS yönteminde 6. madde diğer yöntemlerde 13. maddedir. DS yöntemi dışındaki diğer yöntemlerin kullanılması durumunda teste yer alan 20 madde arasında ayırıcılık gücü en yüksek olan 10. madde, en düşük olan ise 13. madde olarak belirlenmektedir.

Testte yer alan her bir maddenin, madde ile ölçülmek istenen özelliğe sahip olan bireylerle sahip olmayan bireyleri ayırt edebilmesi beklenir. Madde ayırıcılık indeksi olarak tanımlanan bu parametre, ayırıcılık (discriminative) anlamında geçerlik, dolayısıyla yapı geçerliği ile ilgilidir. Madde ayırıcılık indeksinin .40 ve üzerinde olması idealdir. Bununla birlikte .20 ve üzeri değerler de kabul edilebilir düzeyde bir ayırıcılığın varlığına kanıt olabilmektedir (Lord & Novick, 1968; Özçelik, 1981; Baykul, 2000; Tekin, 2007).

Hiç bir kayıp veri yönteminde, ayırıcılık indeksi .20'nin altında olan madde bulunmamaktadır. Madde ayırıcılık indeksi .40 ve üzerinde olan madde sayısı ise DS yönteminde 20, BO yönteminde 18, Dİ, RA, BM, VÇ ve MCMC yöntemlerinde 17, 0 Atama yönteminde 16, diğer yöntemlerde 14'tür.

SBS 2011 Matematik Testi A Kitapçığında yer alan maddelere yönelik olarak, her bir kayıp veri yöntemine göre ayrı ayrı elde edilen madde güvenilirlik katsayıları kestirimleri Tablo 6'da verilmektedir.

Tablo 6. Kayıp veri yöntemlerine göre madde güvenilirlik katsayıları

Madde	Kayıp Veri Yöntemleri											
	DS	0 Atama	BO	SO	YNO	YNM	Dİ	DE	RA	BM	VÇ	MCMC
1	.39	.24	.31	.21	.23	.22	.26	.21	.26	.24	.27	.27
2	.43	.25	.32	.22	.23	.22	.26	.22	.27	.25	.28	.27
3	.39	.24	.39	.20	.23	.22	.27	.20	.30	.29	.32	.31
4	.31	.17	.32	.12	.15	.14	.18	.12	.21	.21	.21	.21
5	.39	.36	.40	.20	.33	.21	.29	.20	.32	.32	.34	.34
6	.22	.14	.17	.13	.13	.13	.13	.13	.13	.12	.14	.13
7	.40	.34	.41	.23	.32	.26	.30	.23	.35	.35	.36	.37
8	.29	.21	.27	.15	.19	.16	.19	.15	.20	.21	.21	.20
9	.43	.30	.35	.27	.29	.27	.31	.27	.32	.30	.33	.32
10	.44	.36	.43	.32	.34	.34	.36	.32	.39	.38	.40	.40
11	.41	.33	.41	.27	.31	.29	.32	.27	.36	.35	.36	.36
12	.43	.29	.37	.25	.27	.26	.30	.25	.32	.31	.33	.33
13	.24	.09	.13	.08	.08	.08	.09	.08	.09	.08	.09	.09
14	.32	.22	.28	.19	.20	.20	.21	.19	.22	.21	.23	.22
15	.41	.34	.35	.31	.32	.32	.31	.31	.35	.33	.35	.35
16	.38	.29	.34	.27	.27	.27	.29	.27	.30	.29	.32	.30
17	.36	.31	.33	.23	.26	.23	.25	.23	.29	.25	.29	.29
18	.37	.18	.32	.15	.16	.16	.21	.15	.24	.21	.25	.24
19	.37	.12	.26	.10	.11	.10	.15	.10	.16	.13	.17	.16
20	.41	.29	.31	.29	.28	.29	.29	.29	.30	.29	.30	.30

Tablo 6'da görüldüğü gibi her bir kayıp veri yönteminde de madde güvenilirlik katsayıları düşüktür. En yüksek kestirimler DS yönteminde (.22 - .44), en düşük kestirimler ise SO ve DE yöntemlerinde (.08 - .32) elde edilmiştir. Her bir kayıp veri yönteminde de güvenilirlik kestirimi en yüksek olan madde 10. maddedir. Güvenirlik kestirimi en düşük olan madde ise DS yönteminde 6. madde, diğer yöntemlerde 13. maddedir.

Madde güvenilirlik katsayısı kestirimleri arasındaki farklara yönelik olarak, bu kestirimler arasında, pozitif yönlü, çoğunlukla orta düzeyde manidar ilişkiler bulunduğu belirlenmiştir (Sperman rho katsayıları .67-.99 arasında ve $p < .01$). Friedman'ın sıra farklarına dayalı ilişkili örneklem çift yönlü varyans analizi sonuçları ($p < .01$) ve Kendall'in ilişkili örneklemler konkordans katsayısı kestirimleri ($p < .01$), genel olarak her bir kayıp veri yöntemi grubuna göre elde edilen kestirimler arasında .01 düzeyinde manidar fark bulunduğunu göstermektedir. Buna göre 0 Atama yöntemi, silmeye dayalı yöntem, basit atamaya dayalı yöntemler ve bir grup olarak MVA-ÇVA yöntemleri ile elde edilen kestirimler arasındaki farklar manidardır.

Test Parametreleri

SBS 2011 Matematik Testi A Kitapçığına yönelik test parametrelerinden toplam puanların dağılımına yönelik betimsel kestirimler ile test ortalama güçlük indeksi, test ortalama ayırıcılık indeksi ve test güvenilirlik katsayıları, her bir kayıp veri yöntemine göre ayrı ayrı kestirilmiş ve aşağıda sunulmuştur.

Farklı kayıp veri yöntemlerine göre elde edilen toplam puanlarının dağılımlarına yönelik betimsel kestirimler Tablo 7'de verilmektedir.

Tablo 7. Kayıp veri yöntemlerine göre toplam puanların dağılımına yönelik betimsel kestirimler

Kayıp Veri Yöntemleri	Min.	Max.	μ	SE (μ)	σ	Basıklık	Çarpıklık
DS	0.0	20.0	8.95	.17	6.26	.63	-1.19
0 Atama	0.0	20.0	6.57	.07	4.87	1.21	.63
BO	0.0	20.0	7.30	.14	5.61	.90	-0.56
SO	0.0	20.0	7.93	.06	4.44	1.08	.42
YNO	0.0	20.0	6.72	.07	4.82	1.20	.61
YNM	0.0	20.0	7.25	.06	4.57	1.23	.71
DI	0.0	20.0	7.93	.06	4.56	.99	.20
DE	0.0	20.0	7.93	.06	4.44	1.08	.42
RA	-2.1	20.0	7.75	.07	4.89	.91	-0.09
BM	0.0	20.0	7.75	.07	4.82	.95	-0.04
VÇ	-5.0	21.0	7.72	.07	4.94	.87	-0.12
MCMC	0.0	20.0	7.77	.07	4.93	.89	-0.18

Tablo 7'de görüldüğü gibi toplam puanlar, RA ve VÇ yöntemleri dışında 0 ile 20 puan aralığında değişmektedir. RA ve VÇ yöntemlerinde kayıp veriler yerine olasılıklı ve ötelemeli atama ile [0, 1] aralığı dışında da değer ataması yapılabilmektedir. DS yönteminde, kayıp veriler yerine değer ataması yapılmamakta, bu veriler silinerek analiz dışı bırakılmaktadır. 0 Atama, YNM, BM ve MCMC yöntemlerinde kayıp veriler yerine 1 ve 0 değerleri atanmaktadır. Diğer yöntemlerde ise [0, 1] aralığında değer ataması yapılmaktadır.

Farklı kayıp veri yöntemlerine göre elde edilen toplam puan ortalamaları, 6.57 ve üzerindedir. En yüksek ortalama DS yönteminde, en düşük ortalama ise 0 Atama

yönteminde elde edilmektedir. 0 Atama yönteminin kullanılması durumunda toplam puan ortalamasının düşmesi, beklenen bir durumdur. DS yönteminde ortalamanın yükselmesi ise ihmal edilen verilerde 0 değerinin 1 değerine göre daha fazla olduğunu göstermektedir. Kayıp veri içeren yanıt örüntülerinde 0 puanlanan yani yanlış yanıtlanan madde sayısının 1 puanlanan yani doğru yanıtlanan madde sayısından yüksek olduğu anlaşılmaktadır. Ortalama kestirimine yönelik standart hata değerleri, .06 ve üzerindedir. En yüksek hatanın DS ve BO yöntemlerinin kullanılması durumunda ortaya çıktığı görülmektedir.

Merkezi yayılma ölçülerinden biri olarak standart sapma değerleri, 4.44 ve üzerindedir. En yüksek standart sapma DS ve BO yöntemlerinde elde edilmektedir. SO ve DE yöntemleri ise en düşük standart hata kestirimlerini üretmektedir.

Basıklık ve çarpıklık katsayılarına göre toplam puanlar sağa çarpık bir dağılıma sahiptir. Test puanlarının dağılımlarının, normal dağılımdan gelmediği iddia edilebilir görülmektedir. Bu bulgu SBS 2011 Matematik Testi A Kitapçığının, izleme testi yapısından çok maksimum performans testi yapısında olduğuna yönelik bir kanıt sağlamaktadır.

Toplam puanların dağılımına yönelik betimsel kestirimlerin dışında diğer test parametreleri Tablo 8'de verilmektedir.

Tablo 8. Kayıp veri yöntemlerine göre test parametreleri

Kayıp Veri Yöntemi	Test Parametreleri				
	Max. Puan (Alt Grup)	Min. Puan (Üst Grup)	Ortalama Güçlük	Ortalama Ayırıcılık	Güvenirlilik (α)
DS	4	15	.447	.636	.923
0 Atama	3	8	.329	.530	.866
BO	3	10	.365	.592	.903
SO	4	8	.363	.502	.844
YNO	4	8	.336	.523	.861
YNM	4	8	.363	.502	.844
DI	5	11	.421	.478	.823
DE	4	8	.363	.502	.844
RA	4	10	.388	.507	.849
BM	4	9	.364	.561	.886
VÇ	4	10	.389	.507	.849
MCMC	4	10	.388	.517	.857

Tablo 8'de görüldüğü gibi toplam gözlem birimlerinin yaklaşık %27'sini oluşturan alt ve üst gruplar, DS yönteminde 39000 civarı gözlem biriminden, MCMC yönteminde 70000 civarı gözlem biriminden, diğer yöntemlerde ise 12000 ile 15000 arasında gözlem biriminden oluşmaktadır. Toplam puanlar sağa çarpık bir dağılım gösterdiği için alt gruplarda yer alan gözlem birimi sayısı, üst gruplara göre daha fazladır.

Alt gruplarda maksimum toplam puanlar 3 ve 5 arasında değişmektedir. En yüksek 'maksimum toplam puan' DI yönteminin kullanılması durumunda, en düşük 'maksimum toplam puan' 0 Atama ve BO yöntemlerinin kullanılması durumunda ortaya çıkmaktadır. Üst gruplarda ise minimum puanlar 8 ile 15 arasında değişmektedir. En yüksek 'minimum toplam puan' DS yönteminde elde edilmektedir. Basit atamaya dayalı kayıp veri yöntemlerinin EÇO ve ÇVA yöntemlerine göre daha düşük kestirimler ürettiği görülmektedir.

Testin ortalama güçlük değerleri kayıp veri yöntemlerine göre .329 ile .447 arasında değişmektedir. En yüksek kestirimler DS ve DI yöntemlerinde, en düşük kestirim 0 Atama

yönteminde elde edilmiştir. Basit atamaya dayalı kayıp veri yöntemlerinin EÇO ve ÇVA yöntemlerine göre genel olarak daha düşük test ortalama güçlüğü kestirimleri ürettiği görülmektedir.

Testin ortalama ayırıcılık indeksi, kayıp veri yöntemlerine göre .502 ile .636 arasında değişmektedir. En yüksek ayırıcılık kestirimleri DS yönteminde, en düşük ayırıcılık kestirimi ise SO, YNM ve DE yöntemlerinde elde edilmektedir. Bu durum DS yönteminin kullanılması durumunda alt ve üst gruplar arasındaki puan farklılıklarının olduğundan daha fazla belirlendiğini göstermektedir. Diğer taraftan SO, YNM ve DE yöntemlerinin kullanılması durumunda alt ve üst gruplar arasındaki puan farklılıkları azalmaktadır.

Test güvenilirliğine yönelik Cronbach'ın α katsayısı kestirimleri, iç tutarlılık anlamında bilgi sağlamaktadır. Kestirilen α katsayıları, kayıp veri yöntemlerine göre .823 ve .923 arasında değişmektedir. Bu durum testin iç tutarlılık anlamında güvenilirliğinin yüksek sayılabilecek düzeyde olduğu yönünde kanıt sağlamaktadır. En yüksek α değeri DS yönteminde, en düşük α değeri ise Dİ, SO, YNM ve DE yöntemlerinde elde edilmiştir. Bu durum silmeye dayalı yöntemlerin, testin iç tutarlılık anlamında güvenilirlik düzeyini olduğundan daha yüksek belirlediğini göstermektedir. Diğer taraftan basit atamaya dayalı yöntemler ise genel olarak testin iç tutarlılık anlamında güvenilirlik düzeyini düşürmektedir.

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu çalışmada SBS 2011 Matematik Testi A Kitapçığına yönelik olarak, farklı kayıp veri yöntemleri ile elde edilen eksiksiz veri setleri üzerinde, klasik test kuramı temelinde madde ve test parametre kestirimleri gerçekleştirilmiştir.

Kullanılan kayıp veri yöntemleri arasında, silmeye dayalı yöntemlerden DS yöntemi en yüksek kestirimleri, 0 Atama yöntemi ise en düşük kestirimleri üretmektedir. Diğer bir deyişle DS yöntemi madde ve test parametrelerini olduğundan daha yüksek, 0 Atama yöntemi ise olduğundan daha düşük kestirme eğilimindedir. DS yönteminin kullanılması durumunda veri setinde 1 puanlanan, 0 Atama yönteminin kullanılması durumunda ise 0 puanlanan madde sayısı artmaktadır. Dolayısıyla her iki yöntem de veri setinin homojenliğinin artmasına yol açmaktadır. Buna bağlı olarak her iki yöntemin de madde güçlük indeksi kestirimlerinde yanlılığa yol açma olasılığının yüksek olduğu değerlendirilmiştir.

Little ve Rubin (1987), Schaffer (1997), Allison (2002, 2009) gibi birçok araştırmacı, özellikle kayıp verilerin tam seçkisizlik özelliğini sağlamadığı ve kayıp veri miktarının yüksek olduğu verilerde, DS gibi silmeye dayalı yöntemlerin yanlı kestirimler ürettiğini belirlemiştir. Bu yöntemler, evrene yönelik kestirimlerde kullanılabilir görülmemektedir. Diğer taraftan Lord (1974), Finch (2008), Hohensin ve Kubinger (2011), Culbertson (2011) gibi birçok araştırmacı, kayıp verilerin 'yanlış' olarak değerlendirilmesine dayalı 0 Atama yönteminin, kayıp verilerin tam seçkisiz ya da kısmî seçkisiz oluşması durumlarında dahi yanlı kestirimlere yol açtığını ve kullanılmasının uygun olmadığını belirlemiştir.

Kayıp verilerle başa çıkmada basit değer atamaya dayalı yöntemler, 0 Atama yöntemi kadar olmamakla birlikte, çoktan seçmeli testlerde genel olarak madde ve test parametrelerini olduğundan daha düşük kestirme eğilimindedir. Bu yöntemlerin yanlılık içermeye olasılıkları, EÇO ve ÇVA yöntemlerine göre daha yüksektir.

Linacre (2004), kayıp verilerin tam seçkisizlik özelliğini sağlamadığı, dolayısıyla ihmal edilebilir olmadığı bu tür durumlarda basit atama yöntemlerinin, ancak EÇO ve ÇVA yöntemleri gibi daha karmaşık yöntemler için başlangıç kestirimlerinin elde edilmesinde kullanılabileceğini belirtmektedir.

Kayıp verilerle başa çıkmada EÇO ve ÇVA yöntemleri, silme ve basit atamaya dayalı yöntemlere göre genel olarak daha 'ılımlı (moderate)' kestirimler üretmektedir. Bununla birlikte VÇ yöntemi, kayıp veriler yerine değer atamada daha geniş bir puan aralığını dikkate alabilmesine bağlı olarak, diğer EÇO ve ÇVA yöntemlerine göre daha yüksek kestirimler üretebilmektedir. MCMC yönteminde ise gerçek veri seti kullanılarak daha büyük bir veri seti üretilmekte ve kestirimler, bu büyük veri seti üzerinde gerçekleştirilmektedir. Bu nedenle MCMC yönteminde parametreler, diğer EÇO ve ÇVA yöntemlerine göre genel olarak olduğundan daha düşük kestirilebilmektedir.

Elashoff ve Elashoff (1971), Lord (1983), Schaffer (1997), DeMars (2002), Acock (2005) gibi birçok araştırmacı, iki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde, özellikle ihmal edilebilirlik sağlanmadığı, yani kayıp verilerin tam seçkisizlik özelliğini sağlamadığı durumlarda, silme ve basit atamaya dayalı yöntemler yerine EÇO yöntemlerinin daha titiz kestirimler ürettiğini belirlemiştir. Rubin (1987), Little ve Rubin (1987) ve Schaffer (1997), bu tür verilerde ÇVA yöntemlerinin de titiz kestirimler ürettiğini belirtmektedir. Kayıp verilerin varlığında EÇO ve ÇVA yöntemlerinin daha iyi performans gösterdiği üzerinde uzlaşıldığı görülmektedir.

Bu çalışmanın sonuçları, tam seçkisizlik özelliğini sağlamayan ancak kısmî seçkisizlik gösteren yani ihmal edilebilir olmayan kayıp verilerin varlığında, çoktan seçmeli testlerde madde ve test parametrelerine yönelik kestirimlerde, uygun bir kayıp veri yönteminin kullanılmasının gerekli olduğunu göstermektedir. Silmeye dayalı yöntemler ve 0 Atama yöntemi, bu tür verilerde kullanılabilir değildir. Basit atama yöntemlerinin ise yanlış kestirimler üretme olasılığı yüksektir. EÇO ve ÇVA yöntemleri, bu tür verilerde kullanılması en uygun kayıp veri yöntemleri olarak değerlendirilmektedir.

Bu çalışmada ulaşılan sonuçlar doğrultusunda, kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan çoktan seçmeli testlerde, madde ve test parametrelerine yönelik kestirimlerde; 0 Atama yönteminin kullanılmaması, silmeye ve basit atamaya dayalı yöntemlerin ancak kayıp verilerin ihmal edilebilir olduğuna yönelik sağlam kanıtlar elde edilebilmişse kullanılması, kayıp veri miktarının sınırlı olduğu ve kayıp veri mekanizması açısından tam seçkisizlik sağlanmasa da kısmî seçkisizliğin sağlandığı durumlarda ise en çok olabilirlik ve çoklu veri atamaya dayalı yöntemlerin tercih edilmesi önerilmektedir.

İleri araştırmalara yönelik olarak; SBS'de yer alan diğer alt test verileri yada benzer yapıdaki başkaca test verileri üzerinde kayıp veri yöntemlerinin performanslarının araştırılması, kayıp veri yöntemlerinin MTK temelinde incelenmesi önerilmektedir.

KAYNAKÇA

- Acock, A. A. (2005). Working with missing values. *Journal of Marriage and Family*, 65, 1012-1028.
- Allison, P. D. (2002). *Missing data*. California: Sage Publication, Inc.
- Allison, P. D. (2009). Missing data. *Quantitative methods in psychology*. (Edt: R. E. Millsao & A. Maydeu-Olivares). London: SAGE Publication. pp. 72-89.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont CA: Wadsworth Thomson Learning Company.

- Culbertson, M. J. (2011). Is it wrong? Handling missing responses in IRT. *Annual Meeting of the National Council on Measurement in Education*, April 2011.
- DeMars, C. (2002). Incomplete data and item parameter estimates under JMLE and MML estimation. *Applied Measurement in Education*, 15, 15-31.
- Demir, E. & Parlak, B. (2012). Türkiye’de eğitim araştırmalarında kayıp veri sorunu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3 (1), 20-241.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39 (B), 1-38.
- Elashoff, J. D. & Elashoff, R. M. (1971). Missing data problems for two samples on a dichotomous variable. *Research and Development Memorandum*, 73, 40-86.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*, 45 (3), 225-245.
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60 (4), 549-576.
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70 (5), 646-675.
- Hohensinn, C. & 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.
- Linacre, J. M. (2004). Rasch model estimation: further topics. *Journal of Applied Measurement*, 5 (1), 95-110.
- Little, R. J. A & Rubin, D. B. (1987). *Statistical analysis with missing data*, 2nd ed. New York: John Wiley & Sons, Inc.
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39, 247-264.
- Lord, F. M. (1983). Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika*, 48, 477-482.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theory of mental test score*. California: Addison-Wesley Publishing.
- Mislevy, R. J. & Wu, P. K. (1996). *Missing responses and IRT ability estimation: omits, choice, time limits, adaptive testing*. ETS Research Report RR-96-30-ONR, Princeton, NJ: Educational Testing Service.
- Özçelik, D. A. (1981). *Okullarda ölçme ve değerlendirme*. Ankara: ÜSYM-Eğitim Yayınları:3.
- Puma, M. J., Olsen, R. B., Bell, S. H. & Price, C. (2009). *What do when data are missing in group randomized controlled trial* (NCE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Educational Sciences, U.S. Department of Education.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (3), 581-592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
- Schaffer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
- Tekin, H. (2007). *Eğitimde ölçme ve değerlendirme* (18. baskı). Ankara: Yargı Yayınları.

Item and Test Parameters Estimations for Multiple Choice Tests in the Presence Of Missing Data: The Case of SBS⁴

Ergül DEMİR⁵

Introduction

Missing data can be defined as the absence of an intended observation. Missing data is very common and widespread when collecting data for research, especially within the behavioral and social sciences. Some participants may miss the questions unconsciously, omit some questions consciously, or just refuse to answer. As another example, the loss of data is frequently seen in longitudinal studies. Some participants die or move from one place to another. Sometimes missing data occurs due to measurement tools or the application itself. Questions or items may not be understandable or readable. The conditions of the application may not be eligible. Due to these reasons, missing data can occur in data sets.

Naturally, it is not possible to make any inference towards the measured trait if there is an absence of an intended observation. Loss of data leads to a range contraction in the data set and a representativeness problem of the sample. In the presence of missing data, estimations may include significant bias and the power of statistical analysis may diminish.

Researchers usually tend to study complete data sets. However, it is often quite difficult to obtain complete data sets. Missing data makes the researchers' job quite difficult. Missing data is an important problem in statistical analysis. Moreover, handling missing data is not easy because of the nature of the standard statistical methods developed in the early 20th century. These still widely-used methods usually need the complete data sets and don't provide any successive solutions for missing data.

Since the 1970s, many methods have been developed for dealing with missing data. These methods can be classified into four categories: methods based on deletion, methods based on simple imputation, methods based on maximum likelihood and methods based on multiple imputation. Methods based on deletion and simple imputation are defined as traditional methods. These traditional methods are now used in much statistical software. Other complicated methods include more complex algorithms and require special software.

Missing data patterns and mechanisms are crucial in determining the appropriate methods of handling missing data problems. Traditional methods based on deletion and simple imputation can be used if the missing data mechanism is defined as completely random. The methods based on maximum likelihood and multiple imputation can be used if the missing data mechanism is defined as completely random or partially random. There is no meaning to use most of the missing data methods if the mechanism is not ignorable. Estimations especially may include significant bias if missing data mechanisms are not met randomly and if there are a lot of missing data in the data set.

Multiple choice tests are usually composed of dichotomously scored items. In this type of categorical data, it is more difficult to deal with the missing data problem. Well-defined prior knowledge for the nature of the lost data is needed. Missing data mechanisms must be examined in detail. An appropriate missing data method has to be determined and used to minimize the possibility of missing data bias.

⁴ This study was produced from some of the data of Ergül Demir's PhD Thesis carried out in Ankara University Educational Sciences Institute under Prof. Dr. Nizamettin Koç's consultancy.

⁵ PhD - Ankara University Faculty of Educational Sciences - erguldemir@ankara.edu.tr

In this study, by using different missing data methods, estimations of item and test parameters for multiple choice tests composed of dichotomously scored items are examined. Its aim is to determine the available and unavailable missing value methods for these kinds of tests.

Methods

This study is a basic research to determine details and relationships. Missing value analyses and parameter estimations were carried out using the SBS 2011 Math Test A Booklet data set. SBS is an achievement test executed by the Turkish Ministry of National Education once a year. This test is for 8th grade students of primary schools. Test results are used for transition to secondary education. SBS is composed of five different sub-tests. The mathematics test is one of these five sub-tests. The data set of the mathematics test contains the response patterns of 527517 participants and some other information with gender and school type and province as independent variables. Also, there are 20 multiple choice items in the test.

Twelve different missing value methods were used to obtain the estimations. The methods are 'listwise deletion' and '0 imputation' and 'serial mean imputation' and 'units mean imputation' and 'mean of nearby points imputation' and 'median of nearby imputation' and 'linear interpolation' and 'linear trend at point' and 'regression imputation' and 'expectation-maximization algorithms' and 'data augmentation' and 'Markov Chain Monte Carlo.' The first of these methods is one of the methods based on deletion. The next seven methods are based on simple imputation. The next three methods are based on the maximum likelihood approach. The last method is based on the multiple imputation approach.

In the analyses, first, the raw data set was re-designed for further analyses. The missing data pattern and mechanism were examined and were attempted to be determined. Second, considering these examinations, the appropriate missing data methods were detected. The complete data sets were obtained for each of the methods separately. On these data sets, 'confirmatory factor analyses (DFA)' were executed. Thus, evidences for the construct validity of the test and the unidimensionality of the test were obtained. Third, the item and test parameters of the test were estimated on each completed data set. Item difficulty index, item discrimination index and item reliability coefficient were estimated as the item parameters. The test average difficulty index, test average discrimination index and test reliability coefficient were estimated as the test parameters. The relationships among the estimated parameters were described and examined.

Findings

Each item in the test includes missing values. Except for one item (the 20th item), the amount of missing value is over five percent for each item. The amount of missing value is over 10 percent for 14 items and over 20 percent for seven items. The missing data pattern shows the structure of the general pattern. The missing data mechanism cannot be defined as completely random. On the other hand, it is highly possible that the missing data mechanism is defined as partially random. As a study of structural validity, DFA results show that items could be defined under one significant factor for each missing data method. Structural validity and unidimensionality could be provided.

Considering item and test parameter estimations, there are middle and high level significant correlations among parameter estimates obtained from different missing data

methods. However, there are significant differences among groups of missing data methods. According to the estimates, methods based on deletion and methods based on simple imputation and, as is the case with one group, methods based on maximum likelihood and multiple imputation significantly differ from one another.

In general, the highest estimates are obtained when using the methods based on deletion. The lowest estimates are obtained when using the methods based on simple imputation. It is highly possible that estimates may include significant bias for both of these groups of missing data methods. More moderate and accurate estimations could be obtained by using the methods based on maximum likelihood and multiple imputation.

Conclusions and Recommendations

Results show that appropriate missing value methods should be determined and used for estimating item and test parameters of multiple choice tests composed of dichotomously scored items in the presence of missing values. When there are lots of missing data, especially over five percent, and when missing data mechanisms cannot be defined as completely random or partially random, methods based on deletion and 0 imputation are not available for this kind of data. Simple imputation methods are likely to produce biased estimates. Methods based on maximum likelihood and multiple imputations are considered to be the most available methods for this kind of data.

For further researches, it is recommended that missing data analyses should be carried out on the other sub-tests of SBS or other similar tests. Also, it can be recommended that the same analyses should be carried out on the basis of 'item response theory.'

Key Words: Missing data, Ignorable, Missing at random, Missing value methods, SBS

Atıf için / Please cite as:

Demir, E. (2013). Kayıp verilerin varlığında çoktan seçmeli testlerde madde ve test parametrelerinin kestirilmesi: SBS örneği [Item and test parameters estimations for multiple choice tests in the presence of missing data: The case of SBS]. *Eğitim Bilimleri Araştırmaları Dergisi - Journal of Educational Sciences Research*, 3 (2), 47–68. <http://ebad-jesr.com/>