International Online Journal of Educational Sciences, 2015, 7 (4), 252 - 265

iojes
www.iojes.net

International Online Journal of Educational Sciences

ISSN: 1309-2707

Comparison of Various Methods Used in Solving Missing Data Problems in terms of Psychometric Features of Scales and Measurement Results under Different Missing Data Conditions

Zekeriya Nartgün[1]

[1]*Abant Izzet Baysal University, Faculty of Education, Bolu, Turkey*

ARTICLE INFO

Article History:
Received 13.08.2015
Received in revised form 04.09.2015
Accepted 10.09.2015
Available online 17.09.2015

ABSTRACT

In this research, five different methods used in solving missing data problems (listwise deletion, series mean, mean of nearby points, multiple imputation and regression imputation) were compared under the conditions of missing completely at random mechanism (MCAR), normal distribution, unidimentionality, different sample sizes (n=150; n=650) and different missing data rates (%5; %10; %20). The comparisons were made within the context of the psychometric features (eigenvalue; explained variance; Cronbach alpha) of the scale used in the study as a data collection tool and the measurement results (normal distribution, mean and standard deviation) obtained as a result of use of the scale. In line with the objective of the study, data were deleted from the complete data sets (n=150; n=650) at different rates (%5; %10; %20) and the missing data sets were transformed into new complete data sets with five different methods used in solving missing data problems. The values obtained from new complete data sets in terms of psychometric features and measurement results were compared with the ones obtained from the complete data sets; and inferences were made on which methods can be more applicable under which conditions compared to the others. For the comparisons, descriptive statistics were used for eigenvalues, explained variance and score distribution; Fisher's z test was used for Cronbach alpha; t test for means and Levene's test for equality of variances for standard deviation. The findings of the research reveal that, under all the conditions within the scope of the research, multiple imputation and regression imputation methods yielded values equal or the nearest to the ones obtained from complete data sets. However, listwise deletion method gave the farest values but the distances are at a negligible level. In comparions made with Fisher's z test, t test and Levene's test for equality of variances, no statistically significant differences were found among the values estimated for missing data methods and complete data sets. Therefore, it was concluded that there are no considerable differences among the methods in terms of applicability to solving the missing data problem.

© 2015 IOJES. All rights reserved

Keywords:
Missing data, psychometric features, test statistics

Extended Summary

Purpose

The purpose of this research is to compare the five different methods used in solving missing data problems (listwise deletion, series mean, mean of nearby points, multiple imputation, and regression imputation) under the conditions of missing completely at random mechanism, normal distribution, unidimentionality, different sample sizes (n=150; n=650) and different missing data rates (%5; %10; %20). The comparisons were made within the context of the psychometric features (eigenvalue; explained variance rate; Cronbach alpha) of the scale used in the study as a data collection tool and the measurement results (normal distribution, mean and standard deviation) obtained as a result of use of the scale.

[1] Corresponding author's address: Abant İzzet Baysal University, Faculty of Education, Department of Educational Sciences, Bolu, Turkey
Telephone: 0 374 254 10 00 / 1670
Fax: 0 374 253 46 41
e-mail: nartgun@yahoo.com
DOI: http://dx.doi.org/10.15345/iojes.2015.04.017

© 2015 International Online Journal of Educational Sciences (IOJES) is a publication of Educational Researches and Publications Association (ERPA)

Method

The study, which is a fundamental research, was conducted based on the data obtained from 650 high school students. The data used in the research were obtained by use of the Academic Dishonesty Tendency Scale developed by Eminoğlu and Nartgün (2009). The scale at a Likert type rating format consists of four subscales and totally 22 items. The height of total score obtained for each subscale and the whole of scale is interpreted as high academic dishonesty tendency.

One of the main conditions in the research is unidimentionality. In order to meet the condition, the research is limited to the data obtained from the second subscale called "tendency towards dishonesty in assignments, and studies such as projects – general" of the scale used in the research. The complete data sets used in the research consist of the data set of 650 students who compeletely answered the 7 items in the subscale and the second data set of 150 students taken from the data set with a random sampling method. Missing data sets were obtained by deleting data at certain rates (%5, %10, %20) at missing completely at random mechanism (MCAR) from the two complete data sets having different sample sizes, in line with the purpose of the study. These data sets with missing data were transformed into new complete data sets with five different methods used in solving missing data problems. In transforming the data sets with missing data into new complete data sets, listwise deletion, series mean, mean of nearby points, multiple imputation and regression imputation methods were used.

In the research, the values obtained from the complete data sets for each different sample sizes and different missing data rates within the context of psychometric features (eigenvalue, explained variance rate; Cronbach alpha) and measurement results (normal distribution, mean, standard deviation) were compared with the data obtained from the new complete data sets. The values obtained from the complete data sets were used as reference values in the comparisons.

The comparison of the values for eigenvalue, explained variance and normality of distribution test was made at descriptive level. Fisher's z test was used to compare Cronbach alpha coefficients; dependent groups t test was used for comparison of means; and Levene's test for equality of variances was used for comparing standard deviations. As a result of the comparisons, inferences were made on which methods can be more applicable, compared to the others, under which conditions.

Results

It was concluded that the values obtained for different missing data rates and different missing data methods within the scope of the research in the context of psychometric features of the scale and the measurement results were similar to the values obtained for complete data sets, despite minor differences. The findings of the research reveal that, under all the conditions within the scope of the research, multiple imputation and regression imputation methods among missing data methods yielded values equal or the nearest to the ones obtained from complete data sets whereas listwise deletion method gave the farest values.

Discussion and Conclusion

The investigations on unidimentionality of the scale were made taking into consideration the size of the difference between eigenvalues of the first and second factor. That the eigenvalue of the first factor is more than three and a half of that of the second factor can be observed in all new data sets composed with application of both the two data sets of 150 and 650 persons and different missing data methods. Thus, it was concluded that each of the missing data methods surveyed in the research had a construct similar to the single factor construct (unidimentional) found for complete data sets. The above-mentioned results are similar to various resarch results in which the construct validity of measurement tools are examined with different missing data rates and different missing data methods (Çokluk and Kayri, 2011; Chen et.al., 2012; Şahin Kürşad, 2014).

In the research, explained variance rate values for both complete data sets and new complete data sets transformed by different missing data methods are higher than 30%, an acceptable rate for a unidimentional scale (Büyüköztürk, 2007), under all the conditions within the scope of the research. In this respect, the applied different missing data methods gave results similar to the ones for the complete data set. On the

other hand, in all different missing data rate conditions, multiple imputation and regression imputation methods gave the nearest values to the ones obtained from complete data sets whereas listwise deletion method gave the farest values. The fact that the explained variance rate values calculated for different missing data methods are lower, despite at a very low level, than the ones obtained from the complete data sets and there was a decrease in all missing data methods depending on the increase at missing data rate of rates at these values is another important result of the study. The results of the study are compatible with the results carried out by Acock, 2005; Hawthorne and Elliot, 2005; Çokluk and Kayri, 2011; Chen et.al., 2012 and Şahin Kürşad, 2014.

As a result of the analysis made with Fisher's z test, in terms of the reliability of the scale, it was concluded that there is no significant difference between the Cronbach alpha realiability coefficients estimated for the complete data sets and the coefficients estimated for new complete data sets transformed by different missing data methods. Meanwhile, the descriptive surveys show that multiple imputation and regression imputation methods resulted in equal or higher coefficients to the coefficients obtained from complete data set while listwise deletion method gave the lowest and far coefficients. This finding is consistent with the studies conducted by Enders (2004). The decrease, despite at a low level, at all missing data methods depending on the increase in the missing data rate at data of calculated coefficients is another important result of the research. These findings are consistent with the results of the studies by Leite and Beretva, 2010; Şahin Kürşad (2014).

Kolmogorov Smirnov Z tests showed that the scale scores distribution of the individuals under the complete data sets is normal. In analysis for the new complete data sets composed as a result of completion of sets which have missing data at different rates with different missing data methods, the scale score distribution of the individuals was found to be normal, as in complete data sets, and there is no difference between the methods in this respect.

In the research, the mean values estimated for both complete data sets and new data sets transformed by different missing data methods, were compared with dependent groups t test under different sample sizes and different missing data rate conditions and it was found there was no statistically significant difference. However, the surveys on descriptive statistics showed that the mean values estimated for different missing data methods were lower than the ones for the complete data sets and the decrease increased as the missing data rate increased, being valid for all missing data methods. The resarch also revealed that listwise deletion method gave the farest results to the values obtained from the complete data set under all the conditions within the scope of the research while multiple imputation and regression imputation methods gave the nearest values. These findings on compariosn of the mean values are consistent with the results of the study by Şahin Kürşad (2014).

In the research, standard deviation values estimated for the complete data sets and new complete data sets transformed by different missing data methods were compared with Levene's test for equality of variances under the different sample sizes and different missing data rate conditions and it found there was no statistically significant difference. The findings on comparison of standard deviation results support the research results in which the standard deviation values calculated for complete data and different missing data methods are similar in a study by Şahin Kürşad (2014).

Based on the results within the conditions of the study, taking into consideration the minor difference among the missing data methods, multiple imputation and regression imputation methods are recommended to be primarily used to solve missing data problems in researches. These are followed by series mean and mean of nearby points in terms of applicability. Listwise deletion is the least recommended method.

Kayıp Veri Sorununun Çözümünde Kullanılan Farklı Yöntemlerin Farklı Kayıp Veri Koşulları altında Ölçeklerin Psikometrik Nitelikleri ve Ölçme Sonuçları Bağlamında Karşılaştırılması

Zekeriya Nartgün[1]

[1]*Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Bolu, Türkiye*

MAKALE BİLGİ

Makale Tarihçesi:
Alındı 13.08.2015
Düzeltilmiş hali alındı 04.09.2015
Kabul edildi 10.09.2015
Çevrimiçi yayınlandı 17.09.2015

ÖZ

raştırmada kayıp veri sorununun çözümünde kullanılan beş farklı yöntem (liste bazında silme, seri ortalaması, yakın noktaların ortalaması, çoklu atama, regresyon ataması) tamamıyla rassal olarak kayıp mekanizması, normal dağılım, tek boyutlu yapı, farklı örneklem büyüklükleri (n=150; n=650) ve farklı kayıp veri oranları (%5; %10; %20) koşulları altında karşılaştırılmıştır. Karşılaştırmalar araştırmada veri toplama aracı olarak kullanılan ölçeğin psikometrik nitelikleri (öz değer; açıklanan varyans; Cronbach alfa) ve ölçeğin kullanılması ile elde edilen ölçme sonuçları (puan dağılımının normalliği, ortalama ve standart sapma) bağlamında gerçekleştirilmiştir. Araştırmanın amacı doğrultusunda öncelikle tam veri setlerinden (n=150; n=650) farklı oranlarda (%5; %10; %20) veri silinmiş, daha sonra bu veriler kayıp veri sorununun çözümünde kullanılan beş farklı yöntemle yeni tam veri setlerine dönüştürülmüştür. Yeni tam veri setlerinden, psikometrik nitelikler ve ölçme sonuçları bağlamında elde edilen değerler tam veri setlerinden elde edilen değerlerle karşılaştırılarak hangi yöntemin hangi koşullar altında diğerlerine göre daha kullanılabilir olduğuna ilişkin çıkarımlarda bulunulmuştur. Karşılaştırmalar öz değerler, açıklanan varyans ve puan dağılımının normalliği için betimsel istatistikler, Cronbach alfa için Fisher'in z testi, ortalama değerleri için t testi ve standart sapma değerleri için Levene'in varyansların homojenliği testi kullanılmak suretiyle gerçekleştirilmiştir. Araştırma bulgularına göre, araştırmada ele alınan tüm koşullar altında, çoklu atama ve regresyon ataması yöntemleri tam veri setlerinden elde edilen değerlere ya eşit ya da en yakın değerleri vermektedir. Liste bazında silme yöntemi ise en uzak değerleri vermekle birlikte bu uzaklıklar ihmal edilebilir düzeydedir. Fisher'in z testi, t testi ve Levene'in varyansların homojenliği testi ile yapılan karşılaştırmalarda tam veri setleri ve farklı kayıp veri yöntemleri ile oluşturulan yeni tam veri setleri için hesaplanan değerler arasında istatistiksel olarak manidar farklılıklar bulunmamıştır. Dolayısıyla araştırmada, kayıp veri sorununun çözümüne yönelik kullanılabilirlikleri bakımından yöntemler arasında büyük benzerliklerin olduğu sonucuna ulaşılmıştır.

© 2015 IOJES. Tüm hakları saklıdır

Anahtar Kelimeler:
Kayıp veri, psikometrik özellikler, test istatistikleri

Giriş

Araştırma sürecinde araştırmacıların karşılaştığı en önemli sorunlardan biri toplanan verilerdeki kayıplardır. Tüm alanlarda karşılaşılmakla birlikte kayıp veri sorunuyla en çok karşılaşılan bilim alanları sosyal bilimler (Vansteelandt ve diğ, 2010) ve davranış bilimleridir (Ginkel ve diğ, 2010). Veri toplama araçlarının uzunluğu sebebiyle katılımcıların bazı maddeleri atlaması, mekanik hatalarla verilerin kaydedilememesi, araştırma içeriğinin cinsellik gibi hassas konuları içermesi (Field, 2009) ve motivasyon eksikliği ya da cevabın bilinememesinden dolayı soruların boş bırakılması (Finch ve Margraf, 2008) verilerdeki kayıpların başlıca sebeplerinden bazılarıdır. Verilerin katılımcılardan kendini ifade etme (self-report) tekniğine dayalı ölçme araçları yoluyla toplandığı durumlarda, kayıp veri sorunu ile daha sık karşılaşılmaktadır (Çokluk ve Kayri, 2011). Veri toplama süreçlerinde katılımcılar veri kaybı yaşanmaması adına bilgilendirilse ya da güdülense dahi yine de verilerde kayıp görülebilmektedir. (Ginkel ve diğ, 2007).

Kayıp veri içeren veri setleri üzerinden gerçekleştirilen analizlerin sonuçlarının yanlılık içerme olasılığı oldukça yüksektir (Van der Ark ve Vermunt, 2010). İstatistiksel analizlerde sıklıkla kullanılan paket programların, analize tabi tutulan verilerin eksiksiz olduğu varsayımına göre hareket ettiği (Bal, 2003) dikkate alındığında, kayıp veri içeren veri setleri ile yapılan analizlerin sonuçlarında yanlışlık ve

[2] Sorumlu yazarın adresi: Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Bolu, Türkiye
Telephone: 0 374 254 10 00 / 1670
Fax: 0 374 253 46 41
e-mail: nartgun@yahoo.com

© 2015 International Online Journal of Educational Sciences (IOJES) is a publication of Educational Researches and Publications Association (ERPA)

yanlılıkların olması kaçınılmazdır. Bu sebeplerle araştırmacıların kayıp veri sorununu ya baştan engellemeleri ya da sorunla karşılaştıkları anda sorunu gidermek için gerekli çözümleri üretmeleri beklenmektedir (McKnight, McKnight, Sidani ve Figueredo, 2007).

1930'lu yıllara dayanmakla beraber kayıp veri sorununun ne anlama geldiği, sebepleri, hangi durumlarda ihmal edilebilir kabul edildiği, sahip olduğu örüntüler, çözüm yolları ve çözüm yollarının istatistiksel paket programlara yansımalarına ilişkin asıl gelişmeler 1980'li yıllarda yaşanmaya başlamıştır (Cheema, 2012; Karal, 2014). Belirtilen tarihlerden günümüze kayıp veri sorununun çözümü için eksik gözlemleri analiz dışı bırakmayı, eksik gözlemler yerine veri atamayı veya çeşitli istatistiksel yöntemlerle eksik verileri tamamlamayı temel alan çok önemli yöntemler geliştirilmiştir (Little, 1988; Duncan ve diğ, 1998; Downey ve King, 1998; Pigott, 2001; McKnight ve diğ, 2007; Ginkel ve diğ, 2007; Young ve diğ, 2011; Van der Ark ve Vermunt, 2011; Carpita ve Manisera, 2011). Liste bazında silme ve çiftler bazında silme gibi eksik verileri analiz dışı bırakmayı temel alan yöntemler araştırmalarda en çok tercih edilenler yöntemler olmuşlardır. Ancak yapılan çalışmalar bu yöntemlerin örneklemde kayba, güvenirlikte azalmaya, tahminlerde yanlılığa neden olduğunu (Oğuzlar, 2001; Satıcı ve Kadılar, 2009; Allison, 2009; Van Der Ark ve Vermunt, 2010; Cumming, 2013) ve yanlılıktan kaynaklı olarak da örneklemin evreni temsil etme derecesinin düştüğünü göstermektedir (Little, 1988; Demir ve Parlak, 2012,). Belirtilen bu sebeplerden dolayı en çok olabilirlik ve çoklu atama temelli "modern" yöntemler geliştirilerek araştırmacıların kullanımına sunulmuşlardır. Bu yöntemler, silme yöntemleri gibi geleneksel olarak adlandırılan yöntemlerin aksine, yanlılığın azaltılması, etkili parametre tahminlerinin yapılması ve daha büyük istatistiksel gücün sağlanması hususunda daha etkili sonuçlar vermeleri sebebiyle araştırmacılara daha çok önerilen yöntemler olmuşlardır (Enders, 2013).

Kayıp veri sorununun çözümünde kullanılan geleneksel ve modern yöntemler çeşitli biçimlerde sınıflandırılabilir. İlgili yöntemleri silme yöntemleri, yaklaşık değer atama yöntemleri, beklenti maksimizasyonu, regresyon ataması ve çoklu atama başlıkları altında ele almak bu sınıflamaya örnek olarak verilebilir. İlgili yöntemlerden bu araştırma kapsamında ele alınanlar incelendiğinde, silme yöntemlerinden biri olan liste bazında silmenin (LBS) bir ya da daha fazla kayıp veri içeren bireylerin veya durumların listeden çıkartılarak sadece tam veri içeren durumların kullanılması esasına dayalı olduğu görülmektedir (Cheema, 2012; Yılmaz, 2014). Yaklaşık değer atama yöntemlerinden seri ortalaması (SO) tüm deneklerin belirli bir değişkene ilişkin ortalaması kayıp verilerin yerine atanmak suretiyle gerçekleştirilmektedir. Yaklaşık değer atama yöntemlerinden yakın noktaların ortalaması (YNO) ise kayıp verinin yakınındaki değerlerin ortalaması alınarak kayıp veri yerine atama yapılması esasına dayanmaktadır (Çokluk ve Kayri, 2011). Çoklu atama (ÇA) yöntemi kayıp veri yerine m tane atamanın yapıldığı tekniktir. Atama sayısı genelde 3-10 arasında değişmektedir ve yöntem atfetme, analiz etme ve bir araya getirme adımlarından oluşmaktadır (Oğuzlar, 2001). Regresyon ataması (RA) ise tam veriler kullanılarak oluşturulan regresyon modeli neticesinde elde edilen değerin kayıp veriler yerine atanması ile gerçekleştirilmektedir (Yılmaz, 2014).

İlgili alanyazın incelendiğinde, kayıp veri sorununu çeşitli biçimlerde ele alan çok sayıda çalışmaya rastlanmaktadır. Örneğin, kayıp veri sorununun çözümünde kullanılan yöntemlerin tanıtıldığı çalışmalar (Little, 1988; Duncan ve diğ, 1998; Downey ve King, 1998; Pigott, 2001; McKnight ve diğ, 2007; Ginkel ve diğ, 2007; Baraldi ve Enders, 2010; Young ve diğ, 2011; Van der Ark ve Vermunt, 2011), araştırmalarda kayıp veri sorununun dikkate alınıp alınmadığının incelendiği çalışmalar (Demir ve Parlak, 2012), hangi bilim dallarında hangi kayıp veri yöntemlerinin tercih edildiğine ilişkin çalışmalar (Pigott, 2001) bu çalışmalara örnek olarak verilebilecek çalışmalardandır. Benzer biçimde kayıp veri sorununun çözümünde kullanılan yöntemlerin hangi koşullar altında nasıl çalıştığına ilişkin karşılaştırmaların yapıldığı çalışmalar (Allison 2003; Bal 2003; Buhi ve diğ, 2008; Allison 2009; Alosh 2009; Baraldi ve Enders, 2010; Çokluk ve Kayri, 2011; Demir 2013; Köse ve Öztemur, 2014; Şahin Kürşad, 2014; Akbaş ve Tavşancıl, 2015) yine konu ile ilgili çalışmalara örnek olarak verilebilecek çalışmalardan bazılarıdır.

Bu araştırmada, kayıp veri sorununun çözümünde sıklıkla kullanılan beş farklı yöntem (liste bazında silme, seri ortalaması, yakın noktaların ortalaması, çoklu atama, regresyon ataması) tamamıyla rassal olarak kayıp mekanizması, normal dağılım, tek boyutlu yapı, farklı örneklem büyüklükleri (n=150; n=650) ve farklı kayıp veri oranları (%5; %10; %20) koşulları altında karşılaştırılmıştır. Karşılaştırmalar araştırmada veri toplama aracı olarak kullanılan ölçeğin psikometrik nitelikleri (öz değer; açıklanan varyans; Cronbach alfa)

ve ölçeğin kullanılması ile elde edilen ölçme sonuçları (puan dağılımının normalliği, ortalama ve standart sapma) bağlamında gerçekleştirilmiştir. Araştırma, yukarıda ifade edilen koşullar altında, kayıp veri sorununun çözümünde kullanılan çeşitli yöntemlerin, gerçek veriler kullanılmak suretiyle, hem ölçeğin psikometrik niteliklerinin hem de ölçme sonuçları bağlamında birlikte ele alınarak incelendiği bir çalışma olması bakımından önemli görülmektedir.

Yöntem

Bu araştırma, kayıp veri sorunun çözümünde kullanılan farklı yöntemlerin farklı kayıp veri koşulları altında, ölçeklerin psikometrik nitelikleri ve ölçme sonuçları bağlamında karşılaştırıldığı bir temel araştırmadır. Temel araştırmalar var olan bilgiye yenilerini eklemek amacıyla gerçekleştirilen teorik veya deneysel nitelikte çalışmalardır (Karasar, 2007).

Çalışma Grubu

Araştırmanın çalışma grubunu 2013-2014 eğitim – öğretim yılında Bolu il merkezinde yer alan liselerde öğrenim gören 650 öğrenci oluşturmaktadır. Öğrencilerin 372'si kız, 278'i ise erkektir. Öğrencilerin sınıflara göre dağılımı ise dokuzuncu sınıftan on ikinci sınıfa doğru sırasıyla 159, 153, 172 ve 166'dır.

Veri Toplama Aracı

Araştırmada kullanılan veriler Eminoğlu ve Nartgün (2009) tarafından geliştirilen Akademik Sahtekârlık Eğilimi Ölçeği ile elde edilmiştir. Açımlayıcı ve doğrulayıcı faktör analizleri ölçeğin dört boyutlu bir yapıya sahip olduğunu göstermiştir. Ölçekte, "kopya çekme eğilimi" adlı birinci alt boyutta 5, "ödev, proje gibi çalışmalarda sahtekârlık eğilimi-genel" adlı ikinci alt boyutta 7, "araştırma yapma ve raporlaştırma sürecinde sahtekârlık eğilimi" adlı üçüncü alt boyutta 4 ve "atıflara yönelik sahtekârlık eğilimi" adlı dördüncü alt boyutta 6 madde yer almaktadır. Her bir alt boyuta ait Cronbach alfa iç tutarlılık güvenirlik katsayıları sırasıyla 0,710; 0,821; 0,785; 0,776'dır. Likert tipi dereceleme ölçeği formatında olan ölçekte "kesinlikle katılıyorum (5)"dan "kesinlikle katılmıyorum (1)"a doğru sıralanan beşli cevap skalası kullanılmıştır. Her bir alt boyut ve ölçeğin bütünü için elde edilen toplam puanın yüksekliği akademik sahtekârlık eğiliminin yüksek olduğunu göstermektedir.

Veriler ve Verilerin Analizi

Araştırmada ele alınan temel koşullardan biri tek boyutluluktur. Bu koşulu sağlamak amacıyla araştırma, araştırmada kullanılan ölçeğin "ödev, proje gibi çalışmalarda sahtekârlık eğilimi-genel" adlı ikinci alt boyutundan elde edilen verilerle sınırlandırılmıştır. Bu alt boyutta yer alan toplam 7 maddeyi eksiksiz cevaplayan 650 öğrenciye ait veri seti ile bu veri setinden tesadüfi örnekleme yöntemi ile çekilen 150 öğrencilik ikinci veri seti araştırmada kullanılan tam veri setlerini oluşturmuşlardır. Farklı örneklem büyüklüklerine sahip olan bu iki tam veri setinden, araştırmanın amacı doğrultusunda, tamamıyla rassal olarak kayıp mekanizması altında, belirli oranlarda (%5, %10, %20) veri silinerek eksik veri setleri oluşturulmuştur. Eksik veri içeren bu setler daha sonra kayıp veri sorununun çözümünde kullanılan beş farklı yöntem ile yeni tam veri setlerine dönüştürülmüştür. Kayıp veri içeren veri setlerinin yeni tam veri setlerine dönüştürülmesinde liste bazında silme, seri ortalaması, yakın noktaların ortalaması, çoklu atama ve regresyon ataması kayıp veri yöntemleri kullanılmıştır.

Araştırmada, farklı örneklem büyüklükleri ve farklı kayıp veri oranlarının her biri için tam veri setlerinden psikometrik nitelikler (öz değer, açıklanan varyans; Cronbach alfa) ve ölçme sonuçları (puan dağılımının normalliği, ortalama, standart sapma) bağlamında elde edilen değerler ile dönüştürülmüş yeni tam veri setlerinden elde edilen değerler karşılaştırılmıştır. Öz değer, açıklanan varyans ve dağılımların normalliği testine ilişkin değerlerin karşılaştırılması betimsel düzeyde yapılmıştır. Cronbach alfa katsayılarının karşılaştırılmasında Fisher'in z testi, ortalamaların karşılaştırılmasında bağımlı gruplar için t testi, standart sapmaların karşılaştırılmasında ise Levene'in varyansların homojenliği testleri kullanılmıştır. Yapılan karşılaştırmalar neticesinde hangi yöntemin hangi durumlarda kayıp veri sorunun çözümünde, diğerlerine göre, daha kullanılabilir olduğuna ilişkin çıkarımlarda bulunulmuştur.

Kayıp veri oranı % 0 olan 150 ve 650 öğrencilik tam veri setlerinden psikometrik nitelikler (öz değer; açıklanan varyans; Cronbach alfa) ve ölçme sonuçları (puan dağılımının normalliği, ortalama ve standart sapma) bağlamında elde edilen değerler aşağıda tablo 1'de verilmiştir. Tam veri setlerinden elde edilen bu

değerler yeni tam veri setlerinden psikometrik nitelikler ve ölçme sonuçları bağlamında elde edilen değerleri karşılaştırarak yorumlamada referans değerler olarak kullanılmıştır.

Tablo 1. Tam veri setleri (n=150; n=650) için hesaplanan öz değer, açıklanan varyans, Cronbach alfa, Kolmogorov Smirnov Z Testi, ortalama ve standart sapma değerleri

Kayıp veri oranı	n	Ö	A.V.%	α	K.S.Z	$\bar{X}$	S
% 0	150	4,87 (1,26)	44,37	0,75	0,685 (0,57)	16,32	3,73
% 0	650	5,12 (1,13)	48,16	0,78	0,523 (0,72)	16,87	3,64

Ö= Öz değer; A.V.= Açıklanan varyans; α= Cronbach alfa; K.S.Z= Kolmogorov Smirnov Z Test; $\bar{X}$= Ortalama; S= Standart sapma

Tablo 1'de görüldüğü üzere, 150 kişilik tam veri seti için hesaplanan öz değerler birinci ve ikinci faktör için sırasıyla 4,87 ve 1,26'dır. Birinci faktöre ait öz değerin ikinci faktöre ait öz değerin üç buçuk katından daha yüksek oluşu ölçülmeye çalışılan yapının tek boyutlu olduğunun bir göstergesidir. Hesaplanan açıklanan varyans değeri, ölçülmeye çalışılan yapıdaki değişkenliğin %44,37'sinin bu tek boyut tarafından açıklandığını göstermektedir. Bu boyuta ait Cronbach alfa iç tutarlılık güvenirlik katsayısı 0,75'dir. Hesaplanan Kolmogorov Smirnov Z testi (z=0,685; p=0,57) bu boyuta yönelik olarak elde edilen ölçme sonuçlarının normal dağılıma sahip olduğunu göstermektedir. Bu boyuta yönelik olarak elde edilen ölçme sonuçlarının ortalama ve standart sapması ise sırasıyla 16,32 ve 3,73'tür.

Yine tablo 1'de görüldüğü üzere 650 kişilik tam veri seti için hesaplanan öz değerler birinci ve ikinci faktör için sırasıyla 5,12 ve 1,13'tür. Birinci faktöre ait öz değerin ikinci faktöre ait öz değerin üç buçuk katından daha yüksek oluşu ölçülmeye çalışılan yapının tek boyutlu olduğunun bir göstergesidir. Hesaplanan açıklanan varyans değeri, ölçülmeye çalışılan yapıdaki değişkenliğin %48,16'sının bu tek boyut tarafından açıklandığını göstermektedir. Bu boyuta ait Cronbach alfa iç tutarlılık güvenirlik katsayısı 0,78'dir. Hesaplanan Kolmogorov Smirnov Z testi (z=0,523; p=0,72) bu boyuta yönelik olarak elde edilen ölçme sonuçlarının normal dağılıma sahip olduğunu göstermektedir. Bu boyuta yönelik olarak elde edilen ölçme sonuçlarının ortalama ve standart sapması ise sırasıyla 16,87 ve 3,64'tür.

Bulgular

150 kişilik tam veri seti, farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için psikometrik nitelikler (öz değer, açıklanan varyans, Cronbach alfa) ve ölçme sonuçları (puan dağılımının normalliği, ortalama, standart sapma) bağlamında elde edilen değerler tablo 2'de verilmiştir.

150 kişilik tam veri seti için hesaplanan öz değerler birinci ve ikinci faktör için sırasıyla 4,87 ve 1,26'dır. Tablo 2 incelendiğinde, farklı kayıp veri oranları için farklı farklı kayıp veri yöntemlerinin uygulanması ile hesaplanan öz değerlerin tam veri setinden elde edilen değerlere oldukça benzer olduğu görülmektedir. Öz değerler bakımından özellikle çoklu atama ve regresyon ataması yöntemlerinin uygulanması ile elde edilen değerler tam veri setinden elde edilenlere en yakın değerler olurken, fark küçük olmakla birlikte liste bazında silme yönteminin uygulanması ile elde edilenler en uzak değerler olmuşlardır. Bir bütün olarak incelendiğinde, uygulanan tüm kayıp veri yöntemlerinin tüm kayıp veri oranlarında, aralarında küçük farklılıklar olmakla birlikte, tam veri seti için belirlenen tek boyutlu yapıya benzer sonuçlar verdiği söylenebilir.

150 kişilik tam veri seti için hesaplanan birinci faktöre ait açıklanan varyans değeri 44,37'dir. Tablo 2 incelendiğinde, farklı kayıp veri oranları ve uygulanan tüm kayıp veri yöntemleri için hesaplanan açıklanan varyans değerlerinin tam veri setinden elde edilene göre düşük olduğu görülmektedir. Bu düşüş kayıp veri oranı arttıkça daha da fazlalaşmaktadır. Tüm farklı kayıp veri oranı koşullarında, çoklu atama ve regresyon ataması yöntemlerinin uygulanması ile elde edilen değerler tam veri setinden elde edilen değere en yakın değerler olurken en uzak değerler liste bazında silme yönteminin uygulanması ile elde edilmişlerdir. Böyle olmakla birlikte, hem tam veri seti için hesaplanan açıklanan varyans değeri hem de farklı kayıp veri yöntemlerinin uygulanması ile elde edilen değerler tek boyutlu bir ölçek için kabul edilebilir olan açıklanan varyans değerinin üzerinde değerler vermişlerdir. Bu yönüyle, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

Tablo 2. 150 kişilik örneklem büyüklüğü, farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için analiz sonuçları

Kayıp veri oranı	YÖNTEM	Ö	A.V. %	α - z	K.S.Z	$\bar{X}$	S
%0		4,87 (1,26)	44,37	0,75	0,685 (0,57)	16,32	3,73
%5	LBS	4,80 (1,32)	38,02	0,72 (0,557)	0,677 (0,59)	16,23	3,68
	SO	4,82 (1,29)	40,55	0,73 (0,377)	0,665 (0,62)	16,26	3,66
	YNO	4,81 (1,29)	41,06	0,73 (0,377)	0,665 (0,62)	16,28	3,66
	ÇA	4,85 (1,26)	43,88	0,77 (-0,403)	0,683 (0,56)	16,30	3,70
	RA	4,85 (1,26)	44,23	0,75 (0,000)	0,683 (0,56)	16,31	3,70
%10	LBS	4,78 (1,33)	37,98	0,71 (0,737)	0,754 (0,48)	16,22	3,64
	SO	4,80 (1,29)	39,05	0,71 (0,737)	0,777 (0,52)	16,24	3,65
	YNO	4,80 (1,30)	39,63	0,72 (0,557)	0,779 (0,52)	16,24	3,66
	ÇA	4,85 (1,27)	44,22	0,76 (-0,197)	0,696 (0,55)	16,29	3,68
	RA	4,84 (1,27)	44,18	0,74 (0,197)	0,698 (0,56)	16,29	3,67
%20	LBS	4,78 (1,36)	38,97	0,70 (0,832)	0,768 (0,49)	16,21	3,64
	SO	4,80 (1,30)	38,01	0,71 (0,737)	0,781 (0,51)	16,24	3,65
	YNO	4,81 (1,30)	38,42	0,71 (0,737)	0,780 (0,51)	16,23	3,65
	ÇA	4,84 (1,27)	44,13	0,75 (0,000)	0,701 (0,56)	16,28	3,70
	RA	4,84 (1,28)	44,10	0,74 (0,197)	0,702 (0,56)	16,28	3,69

LBS= Liste Bazında Silme; SO=Seri Ortalaması; YNO =Yakın Noktaların Ortalaması; ÇA=Çoklu Atama; RA=Regresyon Ataması; Ö=Öz değer; A.V.%=Açıklanan varyans; α=Cronbach alfa; z=Fisher's z testi; K.S.Z=Kolmogorov Smirnov Z Test; $\bar{X}$=Ortalama; S=Standart sapma

150 kişilik tam veri seti için hesaplanan Cronbach alfa güvenirlik katsayısı 0,75 iken farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması ile elde edilen katsayılar 0,70 ile 0,77 arasında değişmektedir. Hesaplanan katsayılar incelendiğinde, çoklu atama ve regresyon ataması yöntemlerinin, farklı kayıp veri oranları için, tam veri setinden elde edilen katsayılara ya eşit ya da en yakın katsayıları verdiği, en düşük ve uzak katsayıların ise liste bazında silme yönteminin uygulanması neticesinde elde edildiği görülmektedir. Elde edilen sonuçlar ayrıca, özellikle liste bazında silme, seri ortalaması ve yakın noktaların ortalaması yöntemleri için Cronbach alfa katsayılarının kayıp veri oranı arttıkça daha düşük değerler aldığını göstermektedir. Böyle olmakla birlikte, Fisher'in z testi ile yapılan analizler tam veri seti için hesaplanan güvenirlik katsayısı ile farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen katsayılar arasında manidar bir farklılığın bulunmadığını ($-1{,}96 \leq z \leq +1{,}96$) göstermektedir. Bu yönüyle, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

Araştırmada ele alınan farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için ölçeğin psikometrik nitelikleri (öz değerler, açıklanan varyans oranları, güvenirlik) bağlamında yapılan analiz sonuçları bir bütün olarak incelendiğinde, ulaşılan sonuçların 150 kişilik tam veri seti için elde edilmiş sonuçlara benzer olduğu söylenebilir. Bir başka deyişle, araştırmada ele alınan koşullar altında, ölçeğin tek boyutlu yapısının korunduğu, ilgili tek boyut tarafından açıklanan varyans değerlerinin yeterince büyük olduğu ve hesaplanan güvenirlik katsayılarının benzer olduğu söylenebilir.

150 kişilik tam veri setini oluşturan bireylerin ölçek puanları normal dağılım göstermektedir (KSZ=0,685; p>0,05). Yapılan Kolmogorov Smirnov Z testleri bireylerin farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen ölçek puanlarının da normal dağıldığını göstermektedir. Bir başka deyişle, araştırmada ele alınan farklı kayıp veri oranları ve yöntemleri koşullar altında, bireylerin ölçek puanları, tam veri setinden elde edilen puanların dağılımına benzer biçimde, normal dağılım göstermektedir. Bu yönüyle, hesaplanan Z ve p değerleri bakımından küçük farklılıklar bulunmakla birlikte, uygulanan kayıp veri yöntemleri arasında farklılık bulunmadığı söylenebilir.

150 kişilik tam veri setini oluşturan bireylerin ölçek puanlarının ortalaması 16,32'dir. Farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen ortalama değerleri ise 16,21 ile 16,31 arasında değişmektedir. Yapılan incelemeler hesaplanan ortalama değerlerin tamamının tam veri setinden elde edilenden daha düşük olduğunu ve bu düşüşün her bir kayıp veri yöntemi için kayıp veri oranı arttıkça daha da fazlalaştığını göstermektedir. Liste bazında silme yönteminin uygulanması ile elde

edilen değerler, tüm farklı kayıp veri oranlarında, tam veri setinden elde edilen değere en uzak değerler olurken en yakın değerleri çoklu atama ve regresyon ataması yöntemleri vermiştir. Böyle olmakla birlikte, yapılan t testleri tam veri setinden elde edilen ortalama değeri ile kayıp veri yöntemlerinin uygulanması neticesinde elde edilen ortalama değerleri arasında manidar bir farklılığın bulunmadığını göstermiştir. Dolayısıyla, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım küçük farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

150 kişilik tam veri setini oluşturan bireylerin ölçek puanlarının standart sapması 3,73'tür. Farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen standart sapma değerleri ise 3,64 ile 3,70 arasında değişmektedir. Yapılan incelemeler hesaplanan ortalama değerlerin tamamının tam veri setinden elde edilenden daha düşük olduğunu ve bu düşüşün her bir kayıp veri yöntemi için kayıp veri oranı arttıkça daha da fazlalaştığını göstermektedir. Liste bazında silme yönteminin uygulanması ile elde edilen değerler, tüm farklı kayıp veri oranlarında, tam veri setinden elde edilen değere en uzak değerler olurken en yakın değerleri çoklu atama ve regresyon ataması yöntemleri vermiştir. Böyle olmakla birlikte, yapılan Levene'in varyansların homojenliği testleri tam veri setinden elde edilen standart sapma değeri ile kayıp veri yöntemlerinin uygulanması neticesinde elde edilen standart sapma değerleri arasında manidar bir farklılığın bulunmadığını göstermiştir. Dolayısıyla, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım küçük farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

Araştırmada ele alınan farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için ölçme sonuçları (puan dağılımının normalliği, ortalama, standart sapma) bağlamında yapılan analiz sonuçları bir bütün olarak incelendiğinde, araştırmada ele alınan koşullar altında, ölçek puanlarının normal dağılım gösterdiği, ortalama ve standart sapma değerlerinin ise manidar farklılık göstermediği görülmüştür. Dolayısıyla, puan dağılımının normalliği, ortalama, standart sapma değerleri bakımından ulaşılan sonuçların 150 kişilik tam veri seti için elde edilmiş sonuçlara benzer olduğu söylenebilir.

650 kişilik tam veri seti, farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için psikometrik nitelikler (öz değer, açıklanan varyans, Cronbach alfa) ve ölçme sonuçları (puan dağılımının normalliği, ortalama, standart sapma) bağlamında elde edilen değerler tablo 3'te verilmiştir.

650 kişilik tam veri seti için hesaplanan öz değerler birinci ve ikinci faktör için sırasıyla 4,98 ve 1,13'tür. Tablo 3'e göre, öz değerler bakımından özellikle çoklu atama ve regresyon ataması yöntemlerinin uygulanması ile elde edilen değerler tam veri setinden elde edilenlere en yakın değerler olurken liste bazında silme yönteminin uygulanması ile elde edilenler en uzak değerler olmuşlardır. Böyle olmakla birlikte bir bütün olarak incelendiğinde, uygulanan tüm kayıp veri yöntemlerinin tüm kayıp veri oranlarında, aralarında küçük farklılıklar olmakla birlikte, tam veri seti için belirlenen tek boyutlu yapıya benzer sonuçlar verdiği söylenebilir.

650 kişilik tam veri seti için hesaplanan birinci faktöre ait açıklanan varyans değeri 48,16'dır. Tablo 3 incelendiğinde, farklı kayıp veri oranları ve uygulanan tüm kayıp veri yöntemleri için hesaplanan açıklanan varyans değerlerinin tam veri setinden elde edilene göre düşük olduğu görülmektedir. Bu düşüş kayıp veri oranı arttıkça daha da fazlalaşmaktadır. Tüm farklı kayıp veri oranı koşullarında, çoklu atama ve regresyon ataması yöntemlerinin uygulanması ile elde edilen değerler tam veri setinden elde edilen değere en yakın değerler olurken en uzak değerler liste bazında silme yönteminin uygulanması ile elde edilmişlerdir. Böyle olmakla birlikte, hem tam veri seti için hesaplanan açıklanan varyans değeri hem de farklı kayıp veri yöntemlerinin uygulanması ile elde edilen değerler tek boyutlu bir ölçek için kabul edilebilir olan açıklanan varyans değerinin üzerinde değerler vermişlerdir. Bu yönüyle, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

Tablo 3. 650 kişilik örneklem büyüklüğü, farklı kayıp veri yöntemleri ve farklı kayıp veri oranları için analiz sonuçları

Kayıp veri oranı	YÖNTEM	Ö	A.V. %	α - z	K.S.Z	$\bar{X}$	S
%0		5,12 (1,13)	48,16	0,78	0,523 (0,72)	16,87	3,64
%5	LBS	4,98 (1,17)	47,97	0,76 (0,883)	0,556 (0,66)	16,74	3,53
	SO	5,03 (1,14)	48,07	0,76 (0,883)	0,589 (0,63)	16,75	3,51
	YNO	5,05 (1,14)	48,07	0,77 (0,450)	0,590 (0,63)	16,75	3,51
	ÇA	5,13 (1,11)	48,10	0,80 (-0,973)	0,526 (0,69)	16,84	3,61
	RA	5,12 (1,11)	48,10	0,79 (-0,468)	0,527 (0,69)	16,83	3,62
%10	LBS	4,94 (1,21)	47,95	0,75 (1,297)	0,638 (0,64)	16,71	3,52
	SO	5,01 (1,15)	48,07	0,76 (0,883)	0,685 (0,61)	16,73	3,46
	YNO	5,02 (1,14)	48,08	0,76 (0,883)	0,682 (0,61)	16,74	3,45
	ÇA	5,14 (1,12)	48,10	0,79 (-0,468)	0,534 (0,69)	16,82	3,60
	RA	5,15 (1,14)	48,10	0,79 (-0,468)	0,530 (0,70)	16,81	3,60
%20	LBS	4,90 (1,24)	47,95	0,75 (1,297)	0,640 (0,64)	16,73	3,49
	SO	4,95 (1,17)	48,06	0,76 (0,883)	0,688 (0,62)	16,75	3,45
	YNO	4,92 (1,15)	48,06	0,76 (0,883)	0,687 (0,61)	16,75	3,47
	ÇA	5,08 (1,11)	48,10	0,78 (0,000)	0,536 (0,71)	16,80	3,60
	RA	5,09 (1,11)	48,10	0,78 (0,000)	0,533 (0,71)	16,81	3,58

LBS= Liste Bazında Silme; SO=Seri Ortalaması; YNO =Yakın Noktaların Ortalaması; ÇA=Çoklu Atama; RA=Regresyon Ataması; Ö=Öz değer; A.V.%=Açıklanan varyans; α=Cronbach alfa; K.S.Z=Kolmogorov Smirnov Z Test; $\bar{X}$=Ortalama; S= Standart sapma

650 kişilik tam veri seti için hesaplanan Cronbach alfa güvenirlik katsayısı 0,78 iken farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması ile elde edilen katsayılar 0,75 ile 0,79 arasında değişmektedir. Hesaplanan katsayılar incelendiğinde, çoklu atama ve regresyon ataması yöntemlerinin, farklı kayıp veri oranları için, tam veri setinden elde edilen katsayılara ya eşit ya da daha yüksek katsayıları verdiği, en düşük ve uzak katsayıların ise liste bazında silme yönteminin uygulanması neticesinde elde edildiği görülmektedir. Böyle olmakla birlikte, Fisher'in z testi ile yapılan analizler tam veri seti için hesaplanan güvenirlik katsayısı ile farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen katsayılar arasında manidar bir farklılığın bulunmadığını ($-1{,}96 \leq z \leq +1{,}96$) göstermektedir. Bu yönüyle, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım farklılıklar olmakla birlikte, güvenirlik katsayıları bakımından tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

Araştırmada ele alınan farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için ölçeğin psikometrik nitelikleri (öz değerler, açıklanan varyans oranları, güvenirlik) bağlamında yapılan analiz sonuçları bir bütün olarak incelendiğinde, ulaşılan sonuçların 650 kişilik tam veri seti için elde edilmiş sonuçlara benzer olduğu söylenebilir. Bir başka deyişle, araştırmada dikkate alınan koşullar altında, ölçeğin tek boyutlu yapısının korunduğu, ilgili tek boyut tarafından açıklanan varyans oranlarının yeterince büyük olduğu ve hesaplanan güvenirlik katsayılarının benzer olduğu söylenebilir.

650 kişilik tam veri setini oluşturan bireylerin ölçek puanları normal dağılım göstermektedir (KSZ=0,532; p>0,72). Yapılan Kolmogorov Smirnov Z testleri bireylerin, farklı kayıp veri oranları için farklı kayıp veri yöntemleri ile elde edilen ölçek puanlarının da normal dağıldığını göstermektedir. Bir başka deyişle, araştırmada ele alınan koşullar altında, bireylerin ölçek puanları, tam veri setinden elde edilen puanların dağılımına benzer biçimde, normal dağılım göstermektedir. Bu yönüyle, hesaplanan Z ve p değerleri bakımından küçük farklılıklar bulunmakla birlikte, uygulanan kayıp veri yöntemleri arasında farklılık bulunmadığı söylenebilir.

650 kişilik tam veri setini oluşturan bireylerin ölçek puanlarının ortalaması 16,87'dir. Farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen ortalama değerleri ise 16,71 ile 16,84 arasında değişmektedir. Elde edilen ortalama değerlerin tamamı tam veri setinden elde edilenden daha düşük olup ve bu düşüş her bir kayıp veri yöntemi için kayıp veri oranı arttıkça daha da fazlalaşmaktadır. Liste bazında silme yönteminin uygulanması ile elde edilen değerler, tüm farklı kayıp veri oranlarında, tam veri setinden elde edilen değere en uzak değerler olurken en yakın değerleri çoklu atama

ve regresyon ataması yöntemleri vermiştir. Böyle olmakla birlikte, yapılan t testleri, tam veri setinden ve farklı kayıp veri yöntemlerinin uygulanması ile oluşturulan yeni veri setlerinden elde edilen ortalama değerleri arasında manidar bir farklılığın bulunmadığını göstermiştir. Dolayısıyla, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım küçük farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

650 kişilik tam veri setini oluşturan bireylerin ölçek puanlarının standart sapması 3,64'tür. Farklı kayıp veri oranları için farklı kayıp veri yöntemlerinin uygulanması neticesinde elde edilen standart sapma değerleri ise 3,45 ile 3,62 arasında değişmektedir. Liste bazında silme yönteminin uygulanması ile elde edilen değerler, tüm farklı kayıp veri oranlarında, tam veri setinden elde edilen değere en uzak değerler olurken en yakın değerleri çoklu atama ve regresyon ataması yöntemleri vermiştir. Böyle olmakla birlikte, yapılan Levene'in varyansların homojenliği testleri tam veri seti ve farklı kayıp veri yöntemlerinin uygulanması neticesinde oluşturulan yeni veri setlerinden elde edilen standart sapma değerleri arasında manidar bir farklılığın bulunmadığını göstermiştir. Dolayısıyla, uygulanan farklı kayıp veri yöntemlerinin, aralarında bir takım küçük farklılıklar olmakla birlikte, tam veri seti için ulaşılan sonuca benzer sonuçlar verdiği söylenebilir.

Araştırmada ele alınan farklı kayıp veri oranları ve farklı kayıp veri yöntemleri için ölçme sonuçları bağlamında yapılan analizler bir bütün olarak incelendiğinde, araştırmada ele alınan koşullar altında, bireylerin ölçek puanlarının normal dağılım gösterdiği, puanlara ait ortalama ve standart sapma değerlerinin ise manidar farklılık göstermediği görülmüştür. Dolayısıyla, puan dağılımının normalliği, ortalama ve standart sapma değerleri bakımından bu çalışmada ele alınan kayıp veri yöntemlerinin 650 kişilik tam veri seti için elde edilmiş sonuçlara benzer sonuçlar verdiği söylenebilir.

Sonuç, Tartışma ve Öneriler

Bu araştırmada kayıp veri sorununun çözümünde kullanılan beş farklı yöntem tamamıyla rassal olarak kayıp mekanizması, normal dağılım, tek boyutlu yapı, farklı örneklem büyüklükleri (n=150; n=650) ve farklı kayıp veri oranları (%5; %10; %20) koşulları altında karşılaştırılmıştır. Karşılaştırmalar veri toplama aracı olarak kullanılan ölçeğin psikometrik nitelikleri (öz değer; açıklanan varyans; Cronbach alfa) ve ölçeğin kullanılması ile elde edilen ölçme sonuçları (puan dağılımının normalliği, ortalama ve standart sapma) bağlamında gerçekleştirilmiştir. Bu amaç doğrultusunda ulaşılan sonuçlar aşağıda verilmiştir.

Ölçeğin tek boyutluluğuna ilişkin incelemeler birinci ve ikinci faktöre ait öz değerler arasındaki farkın büyüklüğü dikkate alınarak yapılmıştır. Birinci faktöre ait öz değerin ikinci faktöre ait öz değerin üç buçuk katından fazla olması hem 150 ve 650 kişilik iki tam veri seti hem de farklı kayıp veri yöntemlerin uygulanması ile oluşturulan yeni veri setlerinin tamamında gözlenen bir durum olmuştur. Dolayısıyla araştırmada ele alınan kayıp veri yöntemlerinin her birinin tam veri setleri için tespit edilen tek boyutlu yapılara benzer bir yapıyı gösterdiği sonucuna ulaşılmıştır. Ulaşılan bu sonuçlar ölçme araçlarının yapı geçerliğinin farklı kayıp veri oranı ve farklı kayıp veri yöntemleri çerçevesinde incelendiği çeşitli araştırma sonuçları ile benzerlikler göstermektedir (Çokluk ve Kayri, 2011; Chen ve diğ, 2012; Şahin Kürşad, 2014).

Araştırmada hem tam veri setleri hem de farklı kayıp veri yöntemleri için hesaplanan açıklanan varyans değerleri, araştırmada ele alınan tüm koşullarda, tek boyutlu bir ölçek için kabul edilebilir oran olan %30 (Büyüköztürk, 2007) değerinin üzerindedir. Bu yönüyle, uygulanan farklı kayıp veri yöntemleri tam veri seti için ulaşılan sonuca benzer sonuçlar vermiştir. Böyle olmakla birlikte tüm farklı kayıp veri oranı koşullarında, çoklu atama ve regresyon ataması yöntemleri tam veri setlerinden elde edilen değerlere en yakın değerleri verirken en uzak değerler liste bazında silme yöntemi vermiştir. Farklı kayıp veri yöntemleri için hesaplanan açıklanan varyans değerlerinin tam veri setlerinden elde edilenlerden az da olsa düşük oluşu ve ayrıca bu değerlerde verilerdeki kayıp veri oranının artışına bağlı olarak, tüm kayıp veri yöntemleri için düşüş gözlenmesi araştırmada ulaşılan bir diğer önemli sonuç olmuştur. Araştırmada ulaşılan bu sonuçlar Roth, 1994; Acock, 2005; Hawthorne ve Elliot, 2005; Baraldi ve Enders, 2010; Çokluk ve Kayri, 2011; Chen ve diğ, 2012 ve Şahin Kürşad, 2014'ın araştırma sonuçları ile tutarlılık göstermektedir.

Fisher's z testi ile yapılan analizler neticesinde, ölçeğin güvenirliği bağlamında, tam veri setleri için hesaplanan Cronbach alfa güvenirlik katsayıları ile farklı kayıp veri yöntemleri için hesaplanan katsayılar arasında manidar bir farklılığın bulunmadığı sonucuna ulaşılmıştır. Böyle olmakla birlikte, betimsel bazda

yapılan incelemeler çoklu atama ve regresyon ataması yöntemlerinin, farklı kayıp veri oranları için, tam veri setinden elde edilen katsayılara ya eşit ya da daha yüksek katsayıları verdiğini, liste bazında silme yönteminin ise en düşük ve uzak katsayıları verdiğini göstermektedir. Bu sonuç Enders (2004) ve Demir (2013)'in çalışmaları ile tutarlılık göstermektedir. Hesaplanan katsayılarda verilerdeki kayıp veri oranının artışına bağlı olarak, tüm kayıp veri yöntemleri için küçük te olsa bir düşüş gözlenmesi araştırmada ulaşılan bir diğer önemli sonuç olmuştur. Araştırmada ulaşılan bu sonuçlar ise Leite ve Beretva, 2010; Çokluk ve Kayri, 2011 ve Şahin Kürşad (2014)'ın araştırma sonuçları ile tutarlılık göstermektedir.

Kolmogorov Smirnov Z testleri tam veri setleri kapsamında yer alan bireylerin ölçek puanlarının dağılımının normal olduğunu göstermiştir. Farklı oranlarda kayıp veri içeren setlerin farklı kayıp veri yöntemleriyle tamamlanması neticesinde oluşan yeni tam veri setleri için yapılan analizlerde, tam veri setlerinde olduğu gibi, bireylerin ölçek puanlarının dağılımlarının normal olduğunu ve bu yönüyle yöntemler arasında farklılıkların bulunmadığını göstermiştir.

Araştırmada, hem tam veri setleri hem de farklı kayıp veri yöntemleri çerçevesinde hesaplanan ortalama değerleri, farklı örneklem büyüklükleri ve farklı kayıp veri oranı koşulları altında, bağımlı gruplar t testi ile karşılaştırılmış ve aralarında istatistiksel olarak manidar bir farklılığın bulunmadığı sonucuna ulaşılmıştır. Ancak betimsel istatistikler çerçevesinde yapılan incelemeler, farklı kayıp veri yöntemleri için hesaplanan ortalama değerlerin tam veri setleri için hesaplanan değerden daha düşük olduğunu ve bu düşüşün, tüm kayıp veri yöntemleri için geçerli olmak üzere, kayıp veri oranı arttıkça daha da fazlalaştığını göstermiştir. Araştırmada ayrıca, liste bazında silme yönteminin, araştırmada ele alınan tüm koşullarda, tam veri setinden elde edilen değere en uzak değerleri verdiği en yakın değerleri ise çoklu atama ve regresyon ataması yöntemlerinin verdiği görülmüştür. Ortalama değerlerinin karşılaştırılmasına yönelik ulaşılan bu sonuçlar Şahin Kürşad (2014)'ın araştırma sonuçlarını destekler niteliktedir.

Araştırmada, hem tam veri setleri hem de farklı kayıp veri yöntemleri çerçevesinde hesaplanan standart sapma değerleri, farklı örneklem büyüklükleri ve farklı kayıp veri oranı koşulları altında, Levene'in varyansların homojenliği testi ile karşılaştırılmış ve aralarında istatistiksel olarak manidar bir farklılığın bulunmadığı sonucuna ulaşılmıştır. Standart sapma değerlerinin karşılaştırılmasına yönelik ulaşılan bu sonuçlar Şahin Kürşad (2014)'ın tam veri ve farklı kayıp veri yöntemleri için hesaplanan standart sapma değerlerinin benzer olduğu araştırma sonuçlarını destekler niteliktedir.

Bu çalışmada ele alınan koşullar çerçevesinde ulaşılan sonuçlar dikkate alındığında, araştırmalarda kayıp veri sorunu ile karşılaşılan durumlarda sorunun çözümü için, yöntemler arasında çok büyük farklılıklar olmamakla birlikte, çoklu atama ve regresyon ataması yöntemlerinin öncelikli olarak kullanılması önerilmektedir. Bu yöntemleri kullanılabilirlik bakımından seri ortalaması ve yakın noktaların ortalaması yöntemleri izlemektedir. Liste bazında silme yöntemi ise kullanılması daha az önerilen yöntemdir.

Kaynakça

Acock, A.C. (2005). Working with missing values. *Journal of Marriage and Family*, 67, 1012-1028.

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.

Allison, P.D. (2003). Missing data techniques for structural equation modeling. *Journal of* A*bnormal Psychology, 112* (4), 545-557, doi: 10.1037/0021-843X.112.4.545.

Allison, P.D. (2009). *Missing data,* London: Sage Publication.

Alosh, M. (2009). The impact of missing data in a generalized integer-valued autoregression model for count data. *Journal of Biopharmaceutical Statistics, 19*(6), 1039-1054, doi: 10.1080/10543400903242787.

Bal, C. (2003). *Çok gruplu veri setlerinde eksik gözlem sorununun çözümlenmesi ve sağlık alanında bir uygulama*. Yayımlanmamış doktora tezi, Eskişehir: Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü.

Baraldi, A.N. ve Enders, C.K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology*, 48, 5–37.

Buhi, E.R., Goodson, P. ve Neilands, T.B. (2008). Out of sight not out of mind: Strategies for handling missing data. *American Journal of Health Behavior*, 32 (1), 83-92.

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

Carpita, M. ve Manisera, M. (2011). On the imputation of missing data in surveys with likert- type scales. *Journal of Classification*, 28, 93-112, doi: 10.1007/s00357-011-9074 z.

Cheema, J. (2012). *Handling missing data in educational research using SPSS*. Unpublished doctoral dissertation, George Mason University, USA.

Chen, S.F., Wang, S. ve Chen, Y.C. (2012). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications*, 39, 4026–4031.

Cumming, P. (2013). Missing data and multiple imputation. *Clinical Review & Education*, 167(7), 656-661.

Çokluk, Ö. ve Kayri, M. (2011). Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenirliği üzerindeki etkisi. *Kuram ve Uygulamada Eğitim Bilimleri*, 11(1), 289-309.

Demir, E. (2013). Kayıp verilerin varlığında çoktan seçmeli testlerde madde ve test parametrelerinin kestirilmesi: SBS örneği. *Eğitim Bilimleri Araştırmaları Dergisi*, 3(2), 47-68.

Demir, E. ve Parlak, B. (2012). Türkiye'de eğitim araştırmalarında kayıp veri sorunu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3(1), 230-241.

Downey, R.G. ve King, C.V. (1998). Missing data in likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, 125(2), 175-191, doi: 10.1080/00221309809595542.

Duncan, T.E., Duncan, S.C. ve Li, F. (1998). A comparison of model - and multiple imputation - based approaches to longitudinal analyses with partial missingness. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(1), 1-21, doi: 10.1080/10705519809540086.

Eminoğlu, E. ve Nartgün, Z. (2009). Üniversite öğrencilerinin akademik sahtekarlık eğilimlerinin ölçülmesine yönelik bir ölçek geliştirme çalışması. *Uluslararası İnsan Bilimleri Dergisi*, 6 (1), 215-240.

Enders, C.K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement*, 64(3), 419-436, doi: 10.1177/0013164403261050.

Enders, C.K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives, 7* (1), 27- 31.

Field, A. (2009). *Discovering statistics using SPSS,* London: Sage Publication.

Finch, H. ve Margraf, M. (2008). *Imputation of categorical missing data: A Comparison of multivariate normal and multinomial methods*. Retrived from http://www.mwsug.org/proceedings/2008/stats/MWSUG-2008-S05.pdf on 05.08.2015.

Ginkel, J.R.V., Van der Ark, L.A., Sijtma, K. ve Vermunt, J.K. (2007). Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. *Computational Statistics & Data Analysis*, 51, 4013-4027, doi:10.1016/j.csda.2006.12.022.

Ginkel, J.R.V., Sijtma, K., Van der Ark, L.A. ve Vermunt, J.K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. *Methodology*, 6(1), 17-30, doi: 10.1027/1614-2241/a000003.

Hawthorne, G. ve Elliot, P. (2005). Imputing cross-sectional missing data: comparison of common techniques. *Australian and New Zealand Journal of Psychiatry*, 39, 583-591, doi:10.1080/j.1440-1614.2005.01630.x.

Karal, Y. (2014). *Cox regresyon yöntemi modelinde kayıp veri analiz yöntemleri*. Yayımlanmamış yüksek lisans tezi, Samsun: Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü.

Karasar, N. (2007). *Bilimsel araştırma yöntemi: kavramlar, ilkeler, teknikler* Ankara: Nobel Yayın Dağıtım.

Köse, İ. A. ve Öztemur, B. (2014). Kayıp veri ele alma yöntemlerinin t-testi ve ANOVA parametreleri üzerine etkisinin incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 400-412.

Leite, W. ve Beretvas, S.N. (2010). The performance of multiple imputation for likert-type items with missing data. *Journal of Modern Applied Statistical Methods*, (9)1, 64-74.

Little, R.J.A. (1988). Missing data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6(3), 287-296.

McKnight, P.E., McKnight, K.M., Sidani, S. Ve Figueredo, A.J (2007). *Missing data: A gentle introduction*. United States of America: The Guilford Press.

Oğuzlar, A. (2001, Eylül). *Alan araştirmalarinda kayip değer problemi ve çözüm önerileri.* V. Ulusal Ekonometri ve İstatistik Sempozyumu'nda sunulan bildiri. Çukurova Üniversitesi, Adana.

Pigott, T.D. (2001). A review of methods for missing data. *Educational Research and Evaluation*, 7(4), 353-383.

Roth, P.L. (1994). Missing data: A conceptual review for applied psychologist. *Personnel Psychology*, 47(3), 37-560.

Satıcı, E. ve Kadılar, C. (2009). Kayıp gözlem olduğunda kitle ortalamasının tahmini. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 10(2), 549-556.

Şahin Kürşad, M. (2014). *Sıklıkla kullanılan kayıp veri yöntemlerinin betimsel istatistik, güvenirlik ve geçerlik açısından karşılaştırılması*. Yayımlanmamış yüksek lisans tezi, Bolu: Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü.

Van der Ark, L. A., ve Vermunt, J. K. (2010). New developments in missing data analysis. *Methodology*, 6(1), 1-2, doi: 10.1027/1614-2241/a000001.

Vansteelandt, S., Carpenter, J. ve Kenward, M.G. (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, 6(1), 37-48. doi: 10.1027/1614-2241/a000005.

Yılmaz, H. (2014). *Random forests yönteminde kayıp veri probleminin incelenmesi ve sağlık alanında bir uygulama*. Yayımlanmamış yüksek lisans tezi, Eskişehir: Eskişehir Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü.

Young, W., Weckman, G. ve Holland, W. (2011) A survey of methodologies for the treatment of missing values within datasets: limitations and benefits, *Theoretical Issues in Ergonomics Science*, 12(1), 15-43, doi: 10.1080/14639220903470205.