



T.C.
EGE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü



**GENELLENEBİLİRLİK KURAMINDA KAYIP VERİ İLE BAŞ
ETME YÖNTEMLERİ ÜZERİNE BİR ARAŞTIRMA**

DİLAY AK
YÜKSEK LİSANS
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
ANABİLİM DALI

İZMİR

2020

T.C.
EGE ÜNİVERSİTESİ
Eđitim Bilimleri Enstitüsü

**GENELLENEBİLİRLİK KURAMINDA KAYIP VERİ
İLE BAŞ ETME YÖNTEMLERİ ÜZERİNE BİR ÇALIŞMA
A RESEARCH ON THE HANDLING MISSING DATA
METHODS IN GENERALIZABILITY THEORY**

DİLAY AK
YÜKSEK LİSANS TEZİ
EĐİTİM BİLİMLERİ ANABİLİM DALI

Tez Danışmanı
DOC. DR. Hakan ATILGAN

İZMİR

2020



EĞİTİM BİLİMLERİ ENSTİTÜSÜ
TEZ SAVUNMA SINAV TUTANAĞI

Dilay AK'ın "Genellenebilirlik Kuramında Kayıp Veri ile Baş Etme Yöntemleri Üzerine Bir Araştırma" başlıklı tezi 17/01/2020 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Ege Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, Eğitim Bilimleri Anabilim Dalı (Eğitimde Ölçme ve Değerlendirme Bilim Dalı) Yüksek Lisans Yeterlik tezi olarak oy birliğiyle kabul edilmiştir.

Prof. Dr. Tuncay ÖĞRETMEN

Jüri Başkanı

Doç. Dr. Hakan ATILGAN (Danışman)

Üye

Dr. Öğr. Üyesi Ahmet Murat ELLEZ

Üye

ETİK KURALLARA UYGUNLUK BEYANI

Ege Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne sunduğum "**Genellenebilirlik Kuramında Kayıp Veri İle Baş Etme Yöntemleri Üzerine Bir Araştırma**" adlı yüksek lisans tezinin tarafımdan bilimsel, ahlak ve normlara uygun bir şekilde hazırlandığını, tezimde yararlandığım kaynakları bibliyografyada ve dipnotlarda gösterdiğimi onurumla doğrularım.

Dilay AK



TEŞEKKÜR

Yüksek lisans eğitimim ve tez yazım sürecinde en yoğun dönemlerinde bile bana zaman ayıran, yardım ve desteğini esirgemeyen, her aşamada yol gösteren değerli danışmanım Sayın Doç. Dr. Hakan ATILGAN'a,

Yüksek lisans eğitimim boyunca, değerli bilgilerini aktaran ve her daim destek veren değerli Ege Üniversitesi Ölçme ve Değerlendirme Anabilim Dalı öğretim üyeleri Sayın Prof. Dr. Tuncay ÖĞRETMEN'e ve Sayın Doç. Dr. Tahsin Oğuz BAŞOKÇU'ya, ayrıca analizlerimde desteğini esirgemeyen Siirt Üniversitesi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Dr. Öğr. Üyesi Sungur GÜREL'e,

Bu süreçte birlikte eğitim aldığım, manevi destek ve motivasyon kaynağı çok değerli arkadaşlarım Yasemin YARDIM, Pelin BAĞDU SÖYLER ve Simge CEYLAN'a,

Yüksek lisans eğitimime başlamamdan bugüne kadar geçen sürede bana her türlü kolaylığı sağlayan Pakmaya Ülkü Hızal Anadolu Lisesi ailesindeki mesai arkadaşlarıma,

Gerek ders dönemi gerekse tez döneminde her zaman yanımda olan ve her zaman desteğini hissettiğim, ihtiyacım olan her zaman yanımda olan ve devam etmemi sağlayan sevgili eşim Ali AK'a, sevgili annem Sunay KAÇICI, babam Berkan KAÇICI ve biricik kardeşim Ahmet Tahsin KAÇICI 'ya,

Eğitim öğretim hayatım boyunca beni yetiştiren tüm öğretmenlerime saygı, şükran ve teşekkürlerimi sunarım.

İÇİNDEKİLER

TEŞEKKÜR.....	i
İÇİNDEKİLER	ii
TABLolar LİSTESİ.....	iii
ŞEKİLLER LİSTESİ	iv
EKLER LİSTESİ	v
KISALTMALAR LİSTESİ.....	vi
ÖZET.....	vii
EXTENDED ABSTRACT	ix
I.GİRİŞ	1
1.1 Problem Cümlesi.....	29
1.1.1 Alt Problemler	29
1.2 Araştırmanın Amacı ve Önemi	31
2. İLGİLİ ARAŞTIRMALAR	32
3. YÖNTEM.....	39
3.1 Araştırmanın Modeli.....	39
3.2 Çalışma Verileri.....	39
3.3 Verilerin Analizi	41
4. BULGULAR	44
4.1 Birinci Alt Probleme Ait Bulgu ve Yorumlar.....	44
4.2 İkinci Alt Probleme Ait Bulgu ve Yorumlar	46
4.3 Üçüncü Alt Probleme Ait Bulgu ve Yorumlar	48
5. SONUÇ, TARTIŞMA ve ÖNERİLER	56
5.1 Uygulayıcılar için öneriler	58
5.2 Araştırmacılar için Öneriler	58
KAYNAKÇA.....	59

TABLÖLAR LİSTESİ

Tablo 1. Tam Veri Seti ve %5 Oranla Oluşturulan Eksik Verili Veri Setlerinden Elde Edilen Evren Puanı Varyansı, Bağlı Hata Varyansı, Mutlak Hata Varyansı, G ve Phi Katsayısı Ortalamaları.....	44
Tablo 2. %5 Oranla Oluşturulan Eksik Verili Veri Setlerinde Hesaplanan Hata İstatistikleri.....	45
Tablo 3. Tam Veri Seti ve %10 Oranla Oluşturulan Eksik Verili Veri Setlerinden Elde Edilen Evren Puanı Varyansı, Bağlı Hata Varyansı, Mutlak Hata Varyansı, G ve Phi Katsayısı Ortalamaları	46
Tablo 4. %10 Oranla Oluşturulan Eksik Verili Veri Setlerinde Hesaplanan Hata İstatistikleri.....	47
Tablo 5. Tam Veri Seti ve %20 Oranla Oluşturulan Eksik Verili Veri Setlerinden Elde Edilen Evren Puanı Varyansı, Bağlı Hata Varyansı, Mutlak Hata Varyansı, G ve Phi Katsayısı Ortalamaları	49
Tablo 6. %20 Oranla Oluşturulan Eksik Verili Veri Setlerinde Hesaplanan Hata İstatistikleri.....	50

ŞEKİLLER LİSTESİ

Şekil 1. Kayıp Veri Örüntüleri	17
Şekil 2. Evren Puanı Varyansına ait RMSE ve Yanlılık Değerleri	51
Şekil 3. Bağlı Hata Varyansına ait RMSE ve Yanlılık Değerleri	52
Şekil 4. Mutlak Hata Varyansına Ait RMSE ve Yanlılık Değerleri	53
Şekil 5. G Katsayısına ait RMSE ve Yanlılık Değerleri	54
Şekil 6. Phi Katsayısına ait RMSE ve Yanlılık Değerleri.....	55



EKLER LİSTESİ

Analizde Kullanılan R Kodları.....	63
------------------------------------	----



KISALTMALAR LİSTESİ

G Kuramı:	Genellenebilirlik Kuramı
G Çalışması:	Genellenebilirlik Çalışması
K Çalışması:	Karar Çalışması
TROK:	Tamamen Rassal Olarak Kayıp
ROK:	Rassal Olarak Kayıp
İEK:	İhmal Edilemez Kayıp
LBS:	Liste Bazında Silme
OA:	Ortalama Atama
RA:	Regresyon Atama
ÇBS:	Çiftler Bazında Atama
BOA:	Bireysel Ortalama Atama
BM:	Beklenti Maksimizasyon
ÇA:	Çoklu Atama
SA:	Sıfır Atama
BA:	Basit Atama
EÇO:	En Çok Olabilirlik
MCMC:	Markov Zinciri Monte Carlo
RMSE:	Hataların Kareleri Ortalamasının Karekökü

ÖZET

GENELLENEBİLİRLİK KURAMINDA KAYIP VERİ İLE BAŞ ETME YÖNTEMLERİ ÜZERİNE BİR ARAŞTIRMA

AK,Dilay

Yüksek Lisans Tezi, Eğitimde Ölçme ve Değerlendirme

Tez Danışmanı: Doç. Dr. Hakan ATILGAN

Ocak, 2020

Bu araştırmada, G kuramında varyans bileşenlerinin ve güvenilirlik katsayılarının kestirilmesi için kullanılan veri setinde kayıp veri olması durumunda; kayıp veri ile baş etme yöntemlerinin karşılaştırmalı olarak incelenmesi amaçlanmaktadır. Bu amaçla 689 kişi, bir test ve dört puanlayıcının olduğu gerçek bir ölçme durumundan elde edilen bir ana veri matrisi kullanılmıştır. Tüm bireyler tüm maddeleri yanıtlamışlar ve tüm puanlayıcılar tüm bireyleri tüm maddelerde puanlamışlardır. Aynı zamanda madde ve puanlayıcı yüzeyleri (facet) random olarak alınmıştır. Bu nedenle veri matrisi çaprazlanmış random bir desen oluşturmuştur. R'da yazılan kodlarla ana veri matrisi %5, %10 ve %20 oranlarında tamamen seçkisiz olarak eksiltilerek normal dağılıma sahip matrisler elde edilmiştir. Ana veri matrisinin özellikleri dikkate alınarak 100 adet veri matrisi simule edilmiştir. Kayıp verili matrisler daha sonra liste bazında silme (LBS), ortalama atama (OA) ve regresyon atama (RA) yöntemleri ile tekrardan tam veri setine dönüştürülmüştür. Verilerin analizinde başlangıçta tam veri setinde G ve K çalışmaları yapılmıştır. G çalışması ile elde edilen varyans bileşenleri kullanılarak G ve Phi katsayıları hesaplanmıştır. Çalışmada ayrıca evren puanı varyansı, mutlak hata varyansı ve bağıl hata varyansı da dikkate alınmış ve yorumlanmıştır. Daha sonra kayıp veri yöntemleri ile tamamlanmış toplamda 100 tekrar üzerinde de aynı analizler yapılmıştır. Her bir kayıp veri oranı ve kayıp veri yöntemi için tekrarlar üzerinden elde edilen G katsayıları, Phi katsayıları, evren puanı varyansları, mutlak hata varyansları ve bağıl hata varyanslarının ortalaması alınmıştır. Farklı oranlarda eksiltilen veri setleri ile tam veri setlerinden elde edilen istatistikler arasındaki sapmalar RMSE (Root Mean Square Error) ve yanlılık ile incelenmiştir.

Araştırmanın sonucunda; LBS yöntemi tüm kayıp veri oranlarında referans değerlere en yakın sonucu vermiştir. OA yöntemi tüm kayıp veri oranlarında evren

puanı varyansı ve mutlak hata varyansı istatistiklerinde farklılaşmanın en fazla olduğu yöntem olmuştur. Bağıl hata varyansı, G katsayısı ve Phi katsayısı için ise RA yöntemi ana veri matrisinden elde edilen istatistiklerden farklılaşmanın en fazla olduğu yöntem olmuştur. Kayıp veri oranının her koşulu için LBS yöntemi kullanılan veri setlerinde düşük hatalı ve yansız kestirimlerin yapıldığı görülmüştür. Evren puanı varyansı ve mutlak hata varyansı istatistiklerinde en yanlış ve yüksek hataya sahip kestirim OA yöntemi kullanılan veri setlerinde elde edilmiştir. Çalışmaya konu olan diğer istatistiklerde en fazla hataya sahip kestirimler RA yöntemi kullanılan verilerde elde edilmiştir. Bağıl hata varyansına ait yanlışlıklar sıfıra çok yakın veya negatif yönlü olarak hesaplanmış ve en yanlış kestirim RA yöntemi kullanılan veri setlerinde elde edilmiştir. G ve Phi katsayısına ait yanlışlık değerlerinde en büyük değer pozitif yönlü olarak RA yönteminde hesaplanmıştır. Tüm istatistiklerde ve kayıp veri baş etme yöntemlerinde kayıp veri oranı arttıkça az da olsa hata değerlerinin arttığı gözlenmiştir.

Anahtar Sözcükler: Kayıp veri ile Baş Etme Yöntemleri, Genellenebilirlik Kuramı

EXTENDED ABSTRACT

A RESEARCH ON THE HANDLING MISSING DATA METHODS IN GENERALIZABILITY THEORY

AK, Dilay

Ms, Department Of Measurement And Evaluation

Advisor: Assoc. Prof. Dr. Hakan ATILGAN

January,2020

Introduction

Measurement has an important role in every aspect of life. It is possible to determine the quantity of many things, express their numerical magnitudes and make decisions in daily life. Measurement is done to determine to what degree an object, an individual or a phenomenon has a certain quality. Determining the properties of objects or phenomena in nature and benefiting from them based on their results are the reasons why measurement is important in scientific studies (Atilgan, Kan, Aydın, 2017; Baykul, 2015; Özçelik, 2016). With the most common definition used in education and psychology measurement is observing any quality and expressing the result of this observation in numbers and symbols (Turgut ve Baykul, 2015).

Accurate results and accurate decisions in scientific studies are possible with valid and reliable measurement instruments. How accurately the scores obtained as a result of the measurement provide information about the characteristics of the individuals or the objects that are subject to the measurement is an important point to be taken into consideration by those who prepare, use and decide the measurement instruments (Atilgan, 2019). Making the right decisions based on the measurement results is possible only when the results are accurate. The error is the difference between the actual value of the property to be measured and the observed value obtained from the measurement. Particularly in social sciences such as education and psychology, there is a high likelihood of interference. The type of error encountered in education is usually a random error. The error mentioned in the classical test theory is a random error (Atilgan, Kan, Aydın, 2017).

Classical test theory is a statistical theory that connects structures in social sciences such as education and psychology with theoretical structure and provides

explanation. The theory basically states that the observed score (X) is the sum of the actual score (T) and the error score (E). The purpose of the measurement instruments is to ensure that the score obtained in terms of the measured property is as close to the actual score as possible. It is possible that the observed score is close to the actual score, with the least error or no error. The degree at which the random error doesn't interfere with the result and measures the desired feature is called the degree of reliability. The closer the results obtained in repeated measurements on the same feature are, the higher the reliability of the measurement results are (Baykul, 2015). Reliability expresses how accurate it is to generalize the observed score of a person to the average score that can be achieved under all possible conditions (Shavelson & Webb, 1991). The difficulty of obtaining repetitive measurements under similar conditions has led researchers to different options to predict reliability. This has revealed different interpretations for reliability. The reliability of the measurement results is named differently depending on the sources of error. Depending on the sources of error; reliability is called stability, consistency, internal consistency and sensitivity. Source of error variance may be the time of measurement, the items or subtests in the measurement instrument, the individual subjected to the measurement or may be something different. Even, the source of error may be the scorer.

The classical test theory considers only one of the potential sources of error mentioned above when considering reliability. It calculates different reliability coefficients for each error source. The sources of error in the measurements made in education and psychology are more than one (Atılgan, 2004). Although the procedures presented in classical test theory are powerful for calculating reliability, they are not sufficient for most measurement situations. The fact that the reliability coefficients obtained by classical test theory which provides only an undifferentiated error component is different from each other has contradictions in terms of interpretation and decision making. This and other limitations of classical test theory have led to the emergence of G theory which evaluates all potential sources of error for a single reliability coefficient using variance analysis (Atılgan, 2019; Crocker & Algina, 2006).

G theory frees and expands the traditional concept of reliability. Reliability examines consistency and inconsistency in observed scores. G theory allows a

researcher to identify and measure sources of inconsistency in observed scores that occur or likely occur on repetitions of a measurement procedure. G theory provides a comprehensive conceptual framework and a strong statistical background for addressing multiple measurement problems. It focuses on the variance components of systematic or non-systematic error sources and their prediction. With the G theory, a single reliability coefficient is calculated by evaluating the error variances resulting from the interaction of potential error sources separately and with each other.

In G theory, non-systematic sources of error other than the individual are called the facet and the levels of these facets are called levels. The concepts include acceptable observations universe and G studies, as well as generalization universe and D (decision) studies. All of the observations that may replace the sample in which the observations are made are called the universe of admissible observation. The universe to be generalized is called the universe of generalization. Individuals or students who are subject to measurement are called the object of measurement (Atılgan, 2005). In G theory; There are two main studies to investigate reliability: Generalizability (G) study and Decision (D) study. With the G study, the variance magnitudes of potential sources of error (item, rater, time, etc.) are found. These scores allow the reliability coefficients (G and Phi) to be calculated (Shavelson and Webb, 1991). The D study uses the information provided by the G study to design the best possible application of the measurement designed for a particular purpose. The aim of the D study is to design an ideal measurement situation by increasing or decreasing the number or levels of error sources such as item, time and rater (Atılgan, 2004; Brennan, 2001; Shavelson & Webb, 1991).

In G theory, random or constant facets have different consequences on the generalizability of measurement. It is called a random facet that the sample size is smaller than the universe size, that the sample is randomly drawn from the universe, or that the sample can be replaced with other samples of the same size taken from the universe. The constant facet is mentioned when the universe is studied in cases where the fixed surface sample size and the universe size are the same, are not randomly selected and are not related to their generalization or cannot be generalized. In G theory, there are patterns determined according to the purpose of measurement, and these patterns differ according to the number and characteristics of

potential sources of error. Data in which one variability resource observes all the conditions of other variability resources is called crossed.

In social sciences such as education and psychology, it is not possible to obtain the same measurement results on the same individuals with the same tool and to make reliable measurements due to random errors involved in the results. G theory eliminates this limitation by generalizing the measurement result obtained from a test to all results that can be obtained. In addition, another factor affecting reliability is the amount of information collected. As the amount of information gathered about a test decreases, the validity of the test result decreases due to the reliability of the score obtained. In other words; In addition to error sources and generalizability to the universe, the amount of data is one of the points that should be considered while working on reliability.

Data are not always fully collected in measuring instruments. At some point, researchers may face the problem of missing quantitative data. Participants may refuse to answer the question or forget, and there may be problems in saving data. While Respondent based situations such as not understanding the question, not taking it seriously, being tired, and health problems may cause missing data; problems apart from the respondent such as equipment failures, data entry errors or data collection problems also play a role in the formation of missing data. Researchers should be cautious when confronted with missing data. When researchers use the missing data methods without careful consideration of the assumptions required by this method, they run the risk of bias and misleading results. (Alpar, 2003; Pigott, 2001; Tabachnick&Fidell, 2001). It is important to decide on how to manage missing data. To overcome this problem, researchers have developed missing datatechniques. The most traditional of these methods are Listwise or Case Data Deletion (LD or CD) and Pairwise Data Deletion (PD) based deletion techniques. Replacement of Missing Observation and Full Observation (Case Substitution), Mean Substitution of Missing Observation, Hot Deck Imputation, Cold Deck Imputation, Regression Imputation, Multiple Imputation and Expectation-Maximization (EM) are the most commonly used methods based on assignment.

In the case of missing data in the data set used while studying the reliability of the psychometric properties of the measurement instrument, some studies are

necessary. The impartiality of the results obtained, the generalizability of the universe and the accuracy of the predictions depend on the extent to which the data loss problem is taken into consideration in the analysis. Taking the missing data into consideration and taking precautions for this are the indication of the fact that the analysis is carried out meticulously. This is true for G theory analysis. Therefore, in case of missing data in the data set used to estimate variance components and reliability coefficients in G theory; missing data techniques should be examined comparatively. The aim of this study is to investigate the handling missing data techniques comparatively in the case of missing data in the data set used to estimate variance components and reliability coefficients in G theory.

Method

In this research, reliability estimation with G theory was performed on data sets completed with different handling missing data techniques. In the study, a master data matrix obtained from a real measurement situation with 689 people, a test and four raters was used. With the codes written in R, the main data matrix was totally randomly decreased by 5%, 10% and 20% and matrices with MCAR mechanism and normal distribution were obtained. In this study, 100 data matrices, in which the characteristics of the main data matrix were taken into account, were simulated. The lost data matrices were then converted back to the complete data set by listwise deletion, mean imputation and regression imputation methods.

In the analysis of the data, G and D studies were performed in the full data set at the beginning. G and Phi coefficients were calculated by using variance components obtained by G study. Universe score variance, absolute error variance and relative error variance were also considered and interpreted in the study. Then, the same analysis was performed on a 100 repetition in total, which was completed with missing data methods. For each missing data rate and missing data method, G coefficients, Phi coefficients, universe score variances, absolute error variances, and relative error variances were averaged over 100 repetitions. The results of the analysis were evaluated by comparing the values obtained from the complete data set and the values obtained from the completed data by the missing data methods. The deviations between the data sets that were subtracted at different rates and the

statistics obtained from the full data sets were examined by RMSE (Root Mean Square Error) and bias.

Results

The RMSE values of universe score variance, relative error variance, absolute error variance, G and Phi coefficients obtained by using listwise deletion method in data sets with 5% missing data vary between .000003 and .000205. The RMSE values of the statistics obtained by using the mean imputation method ranged from .002159 to .141372. Universe score variance was obtained with the highest errors in listwise deletion and mean imputation method. RMSE values were determined between .014955 and .059975 in regression imputation data sets. In this method, it can be said that the most predicted statistic is the G coefficient and the most stable statistic is the relative error variance. It is seen that the bias values of the statistics obtained in the data sets used listwise deletion method vary between -.00235 and .004989. It is regarded that the most biased statistic is relative error variance. The universe score variance, which has the lowest absolute value, was estimated with a bias close to zero. The bias values of the statistics in the datasets using the mean imputation method are negative in all statistics. The variance of the universe score which is the largest in absolute value is the most biased statistic. Phi coefficient has the closest bias to zero. The predicted statistics in the data sets using regression imputation method were estimated with bias ranging between -.04361 and .041961.

RMSE values of universe score variance, relative error variance, absolute error variance, G and Phi coefficients obtained by using listwise deletion method in 10% missing data sets ranged from .000016 to .000514. The universe variance with the highest value is the relative error variance with the lowest value. The RMSE values of the statistics obtained by using the mean imputation method ranged from .006942 to .522555. Universe score variance was estimated with the most errors. RMSE values were determined between .057341 and .229880 in regression imputation data sets. In this method, the absolute error variance has the lowest error value. It can be said that the most predicted statistic has a G coefficient similar to the 5% missing data set. It is seen that the bias values of the statistics obtained in the data sets used listwise imputation method ranged between -.001120 and .009485. It is seen that the most biased statistic is relative error variance. The universe score variance,

which has the lowest absolute value, was estimated with a bias close to zero. The bias values of the statistics in the datasets using the mean imputation method are negative in all statistics. The most biased statistic is the variance of the universe score, which is the largest in absolute value. Phi coefficient has the closest bias to zero. The predicted statistics in the data sets using the regression imputation method were estimated with bias ranging between $-.080570$ and $.075550$. The coefficient G has the smallest bias value by absolute value. The most biased prediction is the absolute error variance which is negative.

RMSE values of universe score variance, relative error variance, absolute error variance, G and Phi coefficients obtained by using listwise deletion method in 20% missing data sets ranged from $.000742$ to $.001726$. The RMSE values of the estimated statistics in the data sets that the mean imputation method was used in vary between $.035139$ and 1.917185 . In the regression imputation method, the RMSE was between $.207503$ and $.822386$. When all the methods are examined; universe score variance and absolute error variance in mean imputation method; relative error variance and G and Phi coefficients were estimated with greater error in regression imputation method. However, listwise deletion method and RMSE values were found to be smaller than the other methods. The bias values of the statistics obtained in the data sets that listwise deletion method was used in ranged between -0.00203 and 0.017321 . It can be said that the most biased statistic is the relative error variance and the least bias is the Phi coefficient. The bias values of the statistics in the datasets that the mean imputation method was used in are negative in all statistics as in the other missing data rates. The largest of the bias values is the variance of the universe score with $-.35417$ and the smallest is the Phi coefficient with $-.02602$. The predicted statistics in the data sets that the regression imputation method was used in were estimated with bias ranging from $-.32938$ to $.164494$. The $.12807$ bias value calculated for the G coefficient has the smallest bias since it is the closest bias value to zero. The most biased prediction is the relative error variance which is negative with $-.32938$, which is farthest from zero.

Conclusion and Discussion

It is seen that the statistics obtained from the missing data method and the statistics obtained from the complete data set differ slightly in all of the missing data

rates. When the statistics obtained from the full data are compared with the statistics obtained from the missing data, the listwise deletion method yields the closest result to all reference data rates.

It was seen that the data sets using listwise deletion method for G and Phi coefficients were similar to the values obtained from the full data set, and as the missing data rate increased, it was found that there were lower values than the full data set coefficients. The mean imputation method was the method with the highest variance in universe score variance and absolute error variance statistics in all missing data rates. The G and Phi coefficients obtained from the data sets using the mean imputation method were found to be lower than the coefficient values obtained from the full data set, and the difference increased slightly as the missing data rate increased. The G and Phi coefficients obtained from the data sets using regression imputation method were found to be higher than the whole data set. For the relative error variance, G coefficient and Phi coefficient, the regression imputation method was the method with the highest difference from the statistics obtained from the main data matrix. It was observed that as the missing data rate increased, differentiation from full data increased but the closest values continued to be obtained by listwise deletion method.

For each condition of the missing data rate, it was observed that low error and stable predictions were made in the data sets that listwise deletion method was used in. Estimation with the highest error in the universe score variance and absolute error variance statistics was obtained from the data sets that the mean imputation method was used in. In the other statistics subject to the study, the estimations with the most errors were obtained from the data using regression imputation method.

As a result of the analysis, listwise deletion method has been the most unbiased estimation in all statistics for each condition of missing data rate. The universe score variance and absolute error variance statistics obtained from the data sets that mean imputation was used in were found to be negative in all conditions of missing data rates and were the most biased estimation. The biases of the relative error variance were calculated as very close to zero or negative direction and the most biased estimation was obtained from the data sets that the regression imputation method was used in. In the bias values of G and Phi coefficient, the highest value

was calculated positively in regression imputation method. When the missing data rate is considered, it can be said that the rate increase in all statistics causes the bias values to grow by absolute value.



I. GİRİŞ

Ölçme, hayatın her alanında önemli bir yere sahiptir. Günlük hayatta birçok şeyin miktarını belirlemek, sayısal büyüklüklerini ifade etmek ve karar verebilmek ölçme ile mümkündür. Günlük hayatın yanı sıra bir bilim dalında çalışmalar yapmak ve uygulamaları ortaya koymak ölçmeyle ve o bilim dalına özgü ölçme araç ve gereçleriyle mümkündür. Ölçme, bir nesnenin, bireyin ya da olgunun bir niteliğe ne derece sahip olduğunu belirlemek amacıyla yapılır. Doğadaki nesne veya olguların sahip oldukları özellikleri saptamak ve sonuçlarına dayanarak onlardan en üst seviyede yarar sağlamak ölçmenin bilimsel çalışmalarda önemli olmasının nedenidir (Atılğan, Kan, Aydın, 2017; Baykul, 2015; Özçelik, 2016). Ölçme için farklı tanımlar yapılmıştır. Campell, ölçmeyi "maddesel değişkenlerin özelliklerine, bu özellikleri düzenleyen kanunlara dayanarak sayılar vermektir." olarak ifade etmiştir (Akt. Baykul, 2015). Magnusson ölçmeyi "geçerli görgül yollarla test edilebilecek kurallar çerçevesinde nesnelere belli özelliklere sahip oluş derecelerine göre sayılar veya semboller vermektir." ifadesiyle tanımlamıştır (Akt. Atılğan, Kan, Aydın, 2017). Eğitim ve psikolojide kullanılan en yaygın tanımıyla ölçme, herhangi bir niteliğin gözlenip gözlem sonucunun sayı ve sembollerle ifade edilmesidir (Turgut ve Baykul, 2015).

Ölçme; kuramsal yapı ve deneysel süreçten oluşan sistem arasındaki ilişkinin kurulması açısından bilimde önemli bir rol alır. Bilimsel bir yapı, tanım, deney ve gözlemlerle kurulur. Yapıyı oluşturan olgular önce tanımlanır sonra deney ve gözlemlerle bulgular elde edilir. Bulguların doğru olup olmadığı kontrol edilir. Doğrulanmayan bulgular hem deneyde hem de yapıda incelenir, gerekirse düzeltilir. Kuramsal ve deneysel çalışmalardaki bu etkileşim bilimin gelişmesinde büyük rol oynar. Deneysel çalışmalarda doğru sonuçların elde edilmesi ve isabetli kararlar verilmesi geçerli ve güvenilir ölçme araçları ile mümkündür. Fiziksel bilimlerde kuramsal ve deneysel yapının kurulması sosyal bilimlere göre daha kolaydır. Diğer bir deyişle nesne veya olguların tanımlanması ve bunların deneysel yolla yoklanması daha kesin sonuçlar verir. Bunun nedeni, sosyal bilimlerin içerdiği değişkenlerin operasyonel olarak tanımlanamaması, ölçülememesi, ölçülebilse bile güvenilir ve geçerli sonuçların elde edilememesidir.

Eğitim ve psikoloji gibi sosyal bilimlerdeki değişkenler örtüktür. Bunlar

yetenek, tutum, başarı, sosyal çevre, problem çözme becerisi vb.dir. Bu değişkenlerin tanımlanmasında ve ölçülmesinde güçlük çekilir. Bu değişkenleri ölçmek için ölçme yöntem ve teknikleri geliştirilmiştir. Eğitimde, eğitim programını değerlendirmek, nitelikli insan gücü yetiştirmek, öğrencilerin gelecekteki eğilimlerini saptamak, bu eğilimlere uygun alanlara yönlendirmek, öğrenme güçlüklerini tespit etmek ve öğretimi iyileştirmek için ölçme aracından elde edilen puanlar kullanılır. Ölçme ile elde edilen puanlara az hatanın karışmış olması nesnel bir değerlendirme için gereklidir. Sonuçları her basamakta kontrol etmek, hataları gidermeden bir sonraki basamağa geçmemek önemlidir. Ölçmede önemli olan özelliğin gerçeğe yakın olarak betimlenmesidir (Baykul, 2015; Özçelik, 2016).

Ölçme sonucu elde edilen puanların birey veya ölçmeye konu olan objelerin özellikleri hakkında ne kadar doğru bilgi sağladığı ölçme araçlarını hazırlayan, kullanan ve karar verenlerin dikkat etmesi gereken önemli bir noktadır (Atılğan, 2019). Hata, ölçülmek istenen özelliğe ait gerçek değer ile ölçme sonucu elde edilen gözlenen değer arasındaki farktır. Ölçme sonuçlarına dayanarak doğru kararlar verebilmek sonuçların hatasız olabilmesi ile mümkündür. Özellikle eğitim ve psikoloji gibi sosyal bilimlerde ölçmeye hata karışma olasılığı yüksektir. Hatayı en aza indirme çalışmaları hata kaynaklarının bulunması ile başlar. Ölçmeye karışan hatalar sabit, sistematik ve tesadüfi olarak üç başlık altında toplanabilir. Sabit hata, ölçmeden ölçmeye değişmeyen hatadır. Sistematik hata, belirli bir duruma göre miktarı değişen hatadır. Hem sistematik hem de sabit hatanın kaynağı, yönü ve miktarı bellidir. Tesadüfi hata ise ölçmeye rastgele karışan kaynağı, yönü ve miktarı bilinmeyen hatadır. Hata; ölçme aracı, ölçmenin yapıldığı ortam, ölçmeyi yapan kişinin dikkatsizliği gibi bir çok durumdan kaynaklanabilir. Eğitimde karşımıza çıkan hata türü genellikle tesadüfi hatadır. Düzeltme yapmak diğer iki hata türünde olduğu kadar kolay değildir. Klasik test kuramında bahsedilen hata tesadüfi hatadır (Atılğan, Kan, Aydın, 2017).

Klasik test kuramı eğitim ve psikoloji gibi sosyal bilimlerde bulunan yapıları kuramsal yapı ile ilişkilendiren ve açıklanmasını sağlayan istatistiksel bir kuramdır. Ölçme, ölçülen özelliğe ait gerçek değeri elde etmek için yapılır. Klasik test kuramında bu gerçek puan olarak isimlendirilir. Ölçme araçlarından elde edilen sonuçlar karışan hata sebebiyle niteliğin gözlenen değeridir. Kuram temel olarak

gözlenen puanın (X), gerçek puan (T) ve hata puanı (E) toplamından oluştuğunu belirtir. Buradan hareketle gözlenen puandaki değişkenlik, gerçek puan varyansı ve hata puanı varyansının toplamına eşittir. Ölçme araçlarında amaç, ölçülen özellik bakımından elde edilen puanın olabildiğince gerçek puana yakın olması sağlamaktır. Ölçme yoluyla elde edilen gözlenen puanlar, gerçek puan ve hatalardan oluşur. Yukarıda tanımlanan gözlenen puandan da anlaşılacağı üzere gözlenen puanın gerçek puana yakın olması hatanın en az olması ya da hatasız olması ile mümkündür. Ölçme sonuçlarına tesadüfi hatanın karışmamış olması güvenilirlik olarak adlandırılır. Aynı özellik üzerinde tekrarlanan ölçmelerde elde edilen sonuçlar birbirine ne kadar yakın ise ölçme sonuçlarının güvenirliliği o kadar yüksektir (Baykul, 2015; Özçelik, 2016). Güvenirlilik, bir testte kişinin gözlemlenen puanının olası tüm koşullar altında elde edebileceği ortalama puana genellemenin ne kadar doğru olduğunu ifade eder. Burada kişinin bilgisinin, tutumunun, becerisinin veya ölçülen diğer nitelik özelliklerinin sabit olduğu varsayılır. Yani, bir kişinin farklı ölçüm durumlarında kazandığı puanlar arasındaki farkların, bir veya daha fazla hata kaynağına bağlı olduğu ve bireyin olgunlaşma veya öğrenme nedeniyle sistematik değişimlerinden kaynaklanmadığı varsayılır. Tek seferde bir puanlayıcı ile belirli bir test formunda elde edilen puan, o kişinin ortalama puanının kabul edilebilir tüm test formları ve puanlayıcı ile eşleşmesinin mümkün olmadığından tamamen güvenilir değildir. Bu kişinin puanı genellikle diğer test formlarında ve puanlayıcılar ile yapılan ölçmelerde farklı olacaktır. Klasik test kuramı, farklı seanslarda yapılan ölçmelerde bahsi geçen test tekrar test güvenirliliği gibi hata kaynaklarını tek seferde ayrı ayrı tahmin edebilir (Shavelson ve Webb, 1991).

Ölçme sorunları bilimin tüm dalları için endişe kaynağıdır. Bilimsel çalışmalar ölçüm prosedürlerine dikkat edilmesini gerektirir. Bilim insanları "ideal" ölçmeyi elde etmenin yollarını ararlar. Ölçüm koşullarını düzeltmek, hatayı azaltmak ve ölçmenin kesinliğini arttırmak için çalışmalar yaparlar. Bu; farklı alanlarda ve farklı nesne özelliklerinde de söz konusudur. Eğitim ve psikolojide tarihsel olarak bu tür meseleler "güvenirlilik" başlığı altında toplanmıştır (Brennan, 2001).

Benzer koşullar altında tekrarlı ölçmeler elde etmenin zorluğu güvenirliliği kestirmek için araştırmacıları farklı seçeneklere yöneltmiştir. Bu da güvenirlilik için farklı anlamlar ortaya çıkarmıştır. Ölçme sonuçlarının güvenirliliği hata kaynaklarına

bağlı olarak farklı isimlendirilir. Test tekrar test yönteminde ölçme puanlarının hata kaynağı zamandır ve güvenilirlik kararlılık olarak isimlendirilir. Paralel testler yönteminde hata kaynağı madde örneklemdir ve güvenilirlik tutarlılık adını alır. İç tutarlılık anlamındaki güvenilirlik ise; ölçme aracındaki maddelerden elde edilen puanların birbiri ile tutarlı olmasıdır. Ölçme amacına uygun olarak yeterli duyarlılıkta olmayan ölçme aracıyla elde edilen sonuçlar daha az güvenilirlerdir. Güvenirlik burada duyarlılık anlamını taşır (Baykul, 2000). Görüldüğü gibi hata varyansının kaynağı; ölçme zamanı, ölçme aracındaki madde ya da alt testler, ölçmeye konu olan birey veya daha farklı olabilir. Bunun yanında hata kaynağı puanlayıcı da olabilir. Birden fazla puanlayıcı test maddelerini puanladığında sonuçlarda farklılıklar olabilir. Hatanın yokluğundan bahsedebilmek için puanların aynı olması gerekir. Bu öznel puanlara dayalı olarak verilen kararların doğruluğunu belirlemek için farklı puanlayıcıların aynı test maddelerine verdikleri puanların birbiriyle tutarlı olması gerekir. Klasik test kuramı güvenilirliği ele alırken yukarıda bahsedilen potansiyel hata kaynaklarından sadece bir tanesini dikkate alır. Her bir hata kaynağı için farklı güvenilirlik katsayısı hesaplar. Eğitim ve psikolojide yapılan ölçmelerde hata kaynakları birden fazladır. Örneğin; paralel testler yönteminde hata kaynağı sadece madde örneklemini olmayabilir. Belki de puanlayıcılar öznel kararlar vererek tutarsız ölçmelere sebep olmuş ve hatada pay sahibi olmuşlardır (Atılğan, 2004).

Klasik test kuramında sunulan prosedürler güvenilirliği hesaplamak için güçlü olsa da çoğu ölçme durumu için yeterli değildir. Bir grup öğrenciye uygulanan test formunu üç puanlayıcının puanladığını varsayalım. Bireylerden elde edilen puanlar iki veya daha fazla olduğunda klasik test kuramındaki istatistiklerin doğrudan uygulanması zorlaşır. Başka bir örnek olarak, bir klinikteki uzmanın özel eğitime ihtiyaç duyan çocuklara uyguladığı alternatif tanı formları klasik test kuramında tanımlandığı şekliyle paralel olmadıkça güvenilirlik katsayısı ve ölçmenin standart hatası yanıltıcı sonuçlar verir. Bir ilaç programında hasta tutumlarının dört puanlayıcı tarafından ve iki farklı seansta ölçüldüğünü varsayalım. Puanlayıcılar arasındaki varyansın ve seanslar arasındaki varyansın bu ölçmedeki hata varyansına ne ölçüde katkıda bulunduğunu ayırmak faydalı olabilir. Sadece farklılaşmamış bir hata bileşeni sağlayan klasik test kuramı, bu ölçme durumlarına kolayca uygulanamaz

(Crocker ve Algina, 2006). Klasik test kuramının bu ve diğer sınırlılıkları Genellenebilirlik (G) kuramına olan ihtiyacı doğurmuştur. Birden çok kaynaktan gelen hataları birlikte analiz etmek ve ortak bir güvenilirlik katsayısı hesaplamak G kuramı ile mümkündür.

G kuramı; güvenilirlik hesaplanması konusunda sınırlılıkları olan kullanım kolaylığı ve işlem basitliği açısından günümüzde hala popüler olan klasik test kuramına tepki olarak ortaya çıkan, varyans analizine dayanan istatistiksel bir kuramdır. Bir ölçme durumunda potansiyel değişkenlik kaynaklarından gelen hatalar güvenilirlik katsayılarını farklı çıkarmak eğilimindedir. Örneğin paralel testler güvenilirliği yüksek olma eğilimindeyken test tekrar test güvenilirliği düşük olabilmektedir. Klasik test kuramı ile elde edilen güvenilirlik katsayılarının birbirinden farklı çıkması yorumlama ve karar verme açısından çelişkiler barındırdığı için varyans analizini kullanarak bütün potansiyel hata kaynaklarını tek bir güvenilirlik katsayısı için değerlendiren G kuramı ortaya atılmıştır. G kuramına ilişkin ilk yayınlar Cronbach, Gieser ve Rajaratnam (1963), Rajaratnam, Cronbach ve Gieser (1965) ve Gieser, Cronbach ve Rajaratnam'ın (1965) makaleleridir. Makalelerin ana katkısı temel olarak yeni güvenilirlik formülleri geliştirmekten ziyade eldeki duruma en uygun olan güvenilirlik katsayısı ve / veya hata varyansını seçmek için prosedürlere yol açan güvenilirlik hakkında bir düşünme tarzının geliştirilmesi olmuştur. Güvenilirlik katsayılarını ve ölçüm hatalarını tahmin etmek için varyans analizinin kullanılması, 1940'lara dayanan bir geçmişe sahiptir. Kuram her ne kadar ilk olarak Cronbach, Gieser ve Rajaratnam (1963) tarafından ortaya atılmış olsa da Hoyt (1941), Lindquist (1953) ve Medley ve Meitzel (1963) gibi yazarlar tarafından da örtük olarak kullanılmıştır. G kuramının olgunlaşması Cronbach, Gieser, Nanda ve Rajaratnam tarafından 1972 yılında yayınlanan "The Dependability of Behavioural Measurement: Theory of Generalizability for Scores and Profile" adlı çalışma ile olmuştur. Shavelson ve Webb 1991 yılında yayınladıkları "Generalizability Theory: A Primer" kitapları ile G kuramını daha anlaşılır hale getirmişlerdir. 2001 yılında Brennan tarafından yayınlanan "Generalizability Theory" kitabıyla hem tek değişkenli hem de çok değişkenli G kuramı detaylı olarak ele alınmıştır (Atılgan, 2019; Crocker ve Algina, 2006).

G Kuramı karar vericinin farklı değerlendirmeler için puanların güvenilirliğini

hesaplamasına olanak sağlar. Klasik test kuramında puanlar hakkında yorum yapabilmek için bağıl değerlendirme kullanılırken, G kuramında buna ek olarak gruptan bağımsız olarak bireyin mutlak performansı hakkında yoruma imkân sağlayan mutlak değerlendirme de kullanılır. G kuramı bu tür yorumlamalar için de puanların güvenilirliği hakkında bilgi verir (Shavelson ve Webb, 1991).

G kuramı geleneksel güvenilirlik kavramını serbestleştirir ve genişletir. Güvenirlik, gözlemlenen puanlardaki tutarlılık ve tutarsızlıkları inceler. G Kuramı, bir araştırmacının, bir ölçüm prosedürünün tekrarları üzerinde ortaya çıkan ya da ortaya çıkabilecek, gözlemlenen puanlardaki tutarsızlık kaynaklarını tanımlamayı ve ölçmeyi sağlar. G kuramı çok sayıda ölçüm sorununu ele almak için kapsamlı bir kavramsal çerçeve ve güçlü bir istatistiksel altyapı sunar. G kuramı bir dereceye kadar ölçme sorunlarına varyans analizi (ANOVA) prosedürlerini uygulayan, klasik test kuramının bir uzantısı olarak görülebilir. Klasik test kuramı, gözlemlenen puanın, gerçek bir puan ve tesadüfi hataya ayrıştırılabileceğini varsaymaktadır. Klasik test kuramı, potansiyel hata kaynaklarını birlikte değerlendirerek eş zamanlı ve tek bir güvenilirlik katsayısı üretmez. G kuramı, klasik test kuramındaki gerçek puanı etkileyen hataya katkıda bulunan birden fazla hata kaynağını birlikte değerlendirmeye olanak sağlayan ANOVA yöntemlerini kullanarak klasik test kuramını serbestleştirir. Sistemik olan ya da sistemik olmayan hata kaynaklarının varyans bileşenlerine ve onların tahminine odaklanır. G kuramı ile potansiyel hata kaynaklarının ayrı ayrı ve birbirleriyle etkileşiminden gelen hata varyansları değerlendirilerek tek bir güvenilirlik katsayısı hesaplanır. Bu güvenilirlik katsayısı genel hatlarıyla klasik test kuramındaki güvenilirlik indeksi ile benzerdir. Güvenirlik indeksi, gerçek puan varyansının gözlenen puan varyansına oranıdır. En genel tanımıyla G kuramı ile elde edilen güvenilirlik katsayısı ise birey varyansının, birey varyansı ve hatalar varyansının toplamına oranıdır. G kuramı sunduğu kavramsal çerçeve ile diğer kuramlardan farklılaşmaktadır (Atılğan, 2004; Brennan, 2001; Shavelson ve Webb, 1991).

Ölçmede potansiyel hata kaynakları; birey, madde, zaman, puanlayıcı vb. olabilir. G kuramında bunlar değişkenlik kaynağı olarak isimlendirilir. Ölçmenin yapıldığı bireyler de varyans kaynağıdır ve bireylerden gelen hata sistemik hatadır. Ölçmenin amacı bireylerin sahip oldukları özelliklerin belirlenmesidir. Birey

dışındaki sistematik olmayan hata kaynakları ise aynı zamanda yüzey (facet) olarak isimlendirilir. Üç öğretmen tarafından puanlanan yirmi maddeden oluşan bir sınav durumunda puanlayıcılar ve maddeler birer yüzeydir. Aynı sınav durumu aralıklarla iki farklı zamanda tekrarlanırsa maddeler, puanlayıcılar ve zaman G kuramında yüzey olarak adlandırılır. Bu yüzeylerin düzeyleri de koşullar (levels) olarak isimlendirilir. Örneğin, maddeler bir yüzey ise birinci madde, ikinci madde,... ,k'ıncı maddenin her biri birer koşuldur. Yukarıdaki üç öğretmen tarafından puanlanan yirmi maddelik sınav durumunda puanlayıcı yüzeyi; birinci puanlayıcı, ikinci puanlayıcı, üçüncü puanlayıcı olarak düzeylere ayrılır. Aynı şekilde madde yüzeyi; birinci madde, ikinci madde, ..., yirminci madde olmak üzere yirmi düzeyden oluşur (Atılğan, 2019). G kuramındaki yüzey ve koşul kavramları varyans analizindeki faktör ve düzey terimleri ile benzerlik gösterir (Brennan, 2001).

G kuramı analizleri sadece çeşitli hata kaynaklarının büyüklüklerini anlamaya değil, aynı zamanda verimli ölçmeler tasarlamaya da yarar. Belki de kuramın en önemli yönü ve benzersiz özelliği kavramsal çerçevesidir. Kavramları arasında kabul edilebilir gözlemler evreni ve G çalışmalarının yanı sıra genelleme evreni ve K (karar) çalışmaları da bulunmaktadır. Gözlemlerin yapıldığı örneklemin yerine geçebilecek gözlemlerin tümüne kabul edilebilir gözlemlerin evreni (the universe of admissible observation) denir. Diğer bir deyişle; yüzeylerin koşullarının tüm olası kombinasyonları olarak açıklanabilir. Genelleme yapılmak istenilen evrene genellenebilirlik evreni (the universe of generalization) denir. Ölçmeye konu olan bireyler veya öğrenciler ise ölçmenin konusu (the object of measurement) olarak adlandırılır (Atılğan, 2005). Ölçülen özelliğe ait madde örneklemini madde evrenini en doğru şekilde temsil etmelidir. Kabul edilebilir tüm maddeler madde evrenine genellenmek istenirse maddeler yüzey olma özelliği taşır (Atılğan, 2019; Crocker & Algina, 2006).

G kuramında; güvenilirliği araştırmak için Genellenebilirlik (G) çalışması ve Karar (K) çalışması olmak üzere iki temel çalışma vardır. G çalışması ile potansiyel hata kaynaklarının (madde, puanlayıcı, zaman vb.) varyans büyüklükleri bulunur. Bu puanlar güvenilirlik katsayılarının (G ve Phi) hesaplanmasına imkân sağlar. G çalışmasının amacı, bir ölçümün birden fazla kullanımını öngörmek ve ölçümdeki varyans kaynakları hakkında en fazla bilgiyi sağlamaktır. Bir G çalışması, mümkün

olduđunca ok sayıda potansiyel varyans kaynađı tanımlamaya ve tasarımına dahil etmeye alıřmalıdır. Bařka bir deyiřle, G alıřmasının kabul edilebilir gzlemlerin evrenini mmkn olduđunca geniř bir řekilde tanımlaması gerekir. G alıřması ayrıca varyans kaynakları hakkında sađladıđı bilgi ile K alıřmasına veri sađlamıř olur (Shavelson ve Webb, 1991).

G alıřması yapan bir arařtırmacı ncelikle genelleme yapılmak istenen evren ile ilgilenir. Evren lme yapılan rneklemedeki kořullardan daha geniř bir dizi kořul barındırır. G alıřması; farklı zamanlarda yapılan lmelerin kararlılıđı, iki veya daha fazla test formunun tutarlılıđı ve alt testler veya maddeler arasındaki iliřki olarak tanımlanmaktadır. K alıřmasında ise sađlanan bilgiler karar vermek iin kullanır. K alıřması ile seme veya yerleřtirme yapmak, grupları karřılařtırmak ve iki veya daha fazla deđiřken arasındaki iliřkiyi bulmak iin bilgi toplanır. G alıřmasının amacı ayrıca yeterli řekilde genelleme yapılabilen K alıřmasının planlanmasına yardımcı olmaktır. G alıřması tasarımı K alıřması iin tm olası tasarımları tahmin etmelidir. Sadece tasarıma dayanarak G alıřması ya da K alıřması olarak sınıflandırma yapmak dođru deđildir. Arařtırmanın amacı belirleyici faktordr. rneđin aynı blgede bulunan farklı okullardaki đrencilere aynı test formunun uygulandıđını dřunelim. Ama her iki okul trndeki đrenciler iin testin gvenirliđin arařtırmak ise yapılan alıřma G alıřmasıdır. Fakat ama, iki đrenci grubunun bařarı seviyelerini karřılařtırmak ve okullar arasındaki eđitim ıktılarını kıyaslamak ise alıřma K alıřması sınıfına girer (Crocker ve Algina, 2006).

K alıřması, belirli bir ama iin tasarlanan lmenin mmkn olan en iyi uygulamasını tasarlamak iin G alıřması tarafından sađlanan bilgileri kullanır. Genelleme yapılmak istenilen evrende kabul edilebilir gzlemlerin deđiřkenlik kaynaklarının dzeyleri belirlenir. Greceli veya mutlak olarak lme hakkında yorum yapılmasını sađlar. K alıřmasının amacı madde, zaman, puanlayıcı gibi hata kaynaklarının birlikte ya da tek tek sayılarının (dzeylerinin) arttırılıp azaltılması ile ideal bir lme durumu tasarlamaktır. Alternatif tasarımlarla hata kaynaklarının varyans byklkleri yorumlanır. Hatayı en aza indirmek, gvenirliđi en st dzeye ıkarmak iin alıřma yapılır. Burada yapılanlar klasik test kuramındaki Sperman-Brown tahmin katsayısında yapılanlar ile benzerlik gsterir. Bir teste madde

eklenmesi ya da puanlayıcı sayısının arttırmak hatayı azaltabilir. Aynı şekilde K çalışmasında da yüzeylerin koşulları değiştirilerek en uygun maliyetli ölçme deseni tasarlanır. G kuramında, klasik test kuramındaki güvenilirlik katsayısına benzer şekilde güvenilirlik seviyesini yansıtan bir özet katsayı bulunur (Atılğan, 2004; Brennan, 2001; Shavelson ve Webb, 1991).

Klasik test kuramında, bireyin gerçek puanı, çok sayıda kesin olarak paralel ölçümlerin ortalaması olarak tanımlanır. Gerçek puan varyansı bu ortalamaların varyansdır ve güvenilirliğin gerçek puanın gözlemlenen puan varyansına oranı olarak tanımlanmıştır. G kuramında, bireyin evren puanı, genelleme evrenindeki ölçümlerin ortalaması olarak tanımlanır. Bu ölçümlerin kesinlikle paralel olduğu varsayılmamaktadır. Bir genellenebilirlik katsayısını tanımlamanın bir yolu, evren puanı varyansının beklenen gözlemlenen varyansa oranıdır. Bu katsayı K çalışmasının tasarımına bağlıdır ve bu nedenle, farklı genellenebilirlik katsayıları K çalışması için farklı tasarımlara bağlıdır. Ayrıca, evren puanı ve dolayısıyla evren puanı varyansı, genellemenin evrenine bağlıdır. Bu nedenle, aynı K çalışma tasarımını kullanmak isteyen ancak farklı genelleme evrenlerine sahip iki araştırmacı, farklı genellenebilirlik katsayıları elde edecektir (Crocker ve Algina, 2006).

G kuramında yüzeylerin tesadüfi veya sabit olması ölçmenin genellenebilirliği üzerinde farklı sonuçlar doğurur. Örneklem boyutunun evren boyutundan küçük olması, örneklemin evrenden rastgele çekilmesi ya da örneklemin evrenden çekilen aynı boyuttaki başka örneklemlemler ile yer değiştirilebilmesi tesadüfi yüzey olarak adlandırılır. Shavelson ve Webb (1991) örneklemin ve yüzeyin tesadüfi olarak nitelendirilebilmesi için "Evrenden alınan koşulların örneklemini aynı büyüklükteki başka koşulların örneklemini ile değiştirmek istiyor muyum?" sorusuna verilen cevabın "evet" olması gerektiğini belirtmiştir. Örneğin; maddelerin yüzey ve her bir maddenin koşul olduğu 40 maddelik bir başarı testinde, maddelerin her biri evrenden rastgele çekilmişse ve evrendeki başka maddelerle değiştirilebiliyorsa madde yüzeyi tesadüfidir. Sabit yüzey örneklem boyutuyla evren boyutunun aynı olduğu, rastgele seçilmeyen ve genellemesi ile ilgilenilmeyen veya genellenilmesi mümkün olmayan durumlarda evren üzerinde çalışıldığında söz konusudur. Matematik, dil bilgisi vb. gibi alt testlerden oluşan başarı testlerinde en küçük

yüzeyleri değiştirmek söz konusu değildir. En küçük yüzeylerin sayısı genelleme evrenindeki yüzeylerin sayısına eşittir. Bu nedenle en küçük yüzey sabit yüzey olarak alınır. Aynı şekilde üç puanlayıcının olduğu bir sınavın yapıldığı okulda eğer puanlayıcı yüzeyinde yer alabilecek başka öğretmen bulunmuyorsa, başka okullardaki öğretmenlere genelleme yapılmayacaksa dolayısıyla evren üzerinde çalışma yapılıyorsa yüzey sabit olarak ele alınır (Atılgan, 2019; Crocker ve Algina, 2006). G kuramında ölçmenin amacına göre belirlenen, potansiyel hata kaynaklarının sayısına ve özelliklerine göre farklılık gösteren desenler vardır. Varyans bileşenlerinin kestirimine yönelik yapılacak analizlere karar verirken hata kaynaklarının sayısından ve özelliklerinden yararlanılır. G çalışması desenine karar vermek için verilerin çaprazlanmış veya yuvalanmış olmalarına bakılır. Çaprazlanmış verilerde değişkenlik kaynaklarının arasına " x " işareti konulur. Bir değişkenlik kaynağının tüm koşulları diğer değişkenlik kaynağının tüm koşullarını gözlemiyorsa bu veri yuvalanmış (nested) olarak adlandırılır. Yuvalanmış desenlerde değişkenlik kaynaklarının arasına ":" işareti konulur. Örneğin, belirli sayıda maddenin yer aldığı bir sınava tabi tutulan bireyler tüm maddeleri cevaplıyor ve hepsi aynı puanlayıcılar tarafından puanlanıyor ise bu çaprazlanmış desendir. Bireyi b , maddeyi m ve puanlayıcıyı p ile gösterilirse bu desen $b \times m \times p$ olarak tanımlanır. Eğer her bir madde farklı bireyler tarafından cevaplanır ve hepsi farklı puanlayıcılar tarafından puanlanır ise bu yuvalanmış desendir ve $b:m:p$ olarak gösterilir. Bireylerin her bir maddeyi cevaplandığı fakat farklı puanlayıcılar tarafından puanlandığı desen ise $m \times (b:p)$ olarak tanımlanır. G çalışmasında veriler tamamen yuvalanmış veya tamamen çaprazlanmış olabileceği gibi yukarıdaki örnekte olduğu gibi hem yuvalanmış hem de çaprazlanmış olabilir (Atılgan, 2004; Shavelson ve Webb, 1991).

Aşağıda birey (b), madde (m) ve puanlayıcı (p) değişkenlik kaynaklarının çaprazlandığı $b \times m \times p$ desenine ait bir G çalışması örneği verilmiştir. G kuramında iki yüzeyli çaprazlanmış desenlerde gözlenen puanların (X_{bmp}) her biri yedi parametreye ayrıştırılarak açıklanır. Desene ait gözlenen puan;

$$\begin{aligned}
 X_{bmp} &= \mu && \text{[Genel Ortalama]} \\
 &+ (\mu_b - \mu) && \text{[Birey Etkisi]} \\
 &+ (\mu_m - \mu) && \text{[Madde Etkisi]}
 \end{aligned}$$

$$\begin{aligned}
& + (\mu_p - \mu) && \text{[Puanlayıcı Etkisi]} \\
& + (\mu_{bm} - \mu_b - \mu_m + \mu) && \text{[Birey-Madde Etkisi]} \quad (1) \\
& + (\mu_{bp} - \mu_b - \mu_p + \mu) && \text{[Birey-Puanlayıcı Etkisi]} \\
& + (\mu_{mp} - \mu_m - \mu_p + \mu) && \text{[Madde-Puanlayıcı Etkisi]} \\
& + (\mu_{bmp} - \mu_{bm} - \mu_{bp} - \mu_{mp} + \mu_b + \mu_m + \mu_p - \mu) && \text{[Kalan Etki]}
\end{aligned}$$

olarak ifade edilmektedir.

Eşitlik 1’de görülebileceği gibi ölçmenin objesi bireye ait gözlenen puan (X_{bmp}); genel ortalama, (1) birey etkisi, (2) madde etkisi, (3) puanlayıcı etkisi, (4) birey - madde ortak etkisi, (5) birey - puanlayıcı ortak etkisi, (6) görev - puanlayıcı ortak etkisi, (7) birey-madde-puanlayıcı ortak etkisi ve artık hata olmak üzere yedi parametreden oluşmaktadır.

Gözlenen puanlara ait varyans;

$$\sigma^2(X_{bmp}) = \sigma^2_b + \sigma^2_m + \sigma^2_p + \sigma^2_{bm} + \sigma^2_{bp} + \sigma^2_{mp} + \sigma^2_{bmp} \quad (2)$$

eşitliği ile gösterilmektedir.

G çalışmasında değişkenlik kaynaklarının ana ve ortak etkileri kestirilebilir. K çalışmasında ise Genellenebilirlik (G) ve Phi (ϕ) olmak üzere iki güvenilirlik katsayısı hesaplanabilir. Hesaplanan katsayılar ölçmenin amacına göre farklılık gösterir. G katsayısı bağıl değerlendirmeler, Phi katsayısı mutlak değerlendirmeler içindir. Bu iki katsayının farkı, G katsayısında bağıl hata varyansı, Phi katsayısında ise mutlak hata varyansının kullanılmasından kaynaklanır.

Mutlak hata ve bağıl hata G kuramında en sık karşılaşılan iki tip hatadır. Araştırmacı bireylerin test performanslarına göre göreceli sıralamasına veya karşılaştırmalı kararlar almak için ölçüm prosedürünün yeterliliğine odaklanabilir. Bu tür kararlar sıklıkla norm referanslı veya test puanlarının göreceli yorumlarıyla ilişkilendirilir. Bu yönde kararlar alabilmek için en önemli nokta, ölçmeye konu olan bireyin puanının izole değil, grup performansı ile ilgili olarak yorumlanmasıdır. Grup performansının en belirgin ölçüsü grup ortalama puanıdır ve ilişkili hataya genellenebilirlik teorisinde "bağıl" hata denir. Bağıl hata varyansı bireyin gözlenen sapma puanı ile evren sapma puanı arasındaki fark olarak tanımlanır ve $\sigma^2(\delta)$ ile

gösterilir. Bağlı hata varyansı klasik kuramdaki hata varyansı ile benzerlik gösterir. Mutlak hata varyansı bireyin gözlenen ortalama puanı ile evren puanı arasındaki farktır ve $\sigma^2(\Delta)$ ile gösterilir (Brennan, 2001).

Varyans bileşenlerinin çeşitli kombinasyonlarla alınması hata varyanslarını farklılaştırmıştır. Bağlı hata varyansında bireyin görece durumunu etkileyen tüm varyans bileşenleri hataya katkıda bulunur. Bu bileşenler, ölçmenin amacı olan bireyin her bir yüzey ile olan ortak etkileşimidir. Mutlak hata varyansı birey hariç tüm varyans bileşenlerinin toplamıdır. Bu bileşenler, tüm etkileşimleri ve yüzey ana etkilerini içerir. Bağlı hata varyansının mutlak hata varyansına göre daha az varyans bileşeni içermesi, bireylerin puanları ile ilgili görece yorumların mutlak yorumlara göre daha az hataya eğilimli olmasına sebep olur (Brennan, 2001; Shavelson ve Webb, 1991).

Bağlı hata varyansı birey ve diğer yüzeylerin ortak etki varyanslarının bileşenlerinden (σ^2_{bm} , σ^2_{bp} ve $\sigma^2_{bmp,e}$) oluşur. Bu sebeple bağlı hata varyansı Eşitlik 3'teki gibi elde edilir.

$$\sigma^2(\delta) = \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{bmp,e}}{n_m n_p} \quad (3)$$

Mutlak hata varyansı birey hariç tüm yüzeylerin ortak ve ana etki varyanslarını kapsar. İki yüzeyli çaprazlanmış desende elde edilen birey dışındaki varyans bileşenlerinin (σ^2_m , σ^2_p , σ^2_{bm} , σ^2_{bp} , σ^2_{mp} ve σ^2_{bmp}) toplamından oluşur. Mutlak hata varyansı Eşitlik 4'teki gibi gösterilir.

$$\sigma^2(\Delta) = \frac{\sigma^2_m}{n_m} + \frac{\sigma^2_p}{n_p} + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{mp}}{n_m \cdot n_p} + \frac{\sigma^2_{bmp,e}}{n_m \cdot n_p} \quad (4)$$

(Atılğan, 2019).

G kuramı her ne kadar varyans bileşenlerinin ve ölçme hatasının önemini vurgulasa da klasik kuramdaki güvenilirlik katsayısına benzer bir genellenebilirlik katsayısı ortaya koyar. Klasik kuramda gözlenen puanlar gerçek puan ve hata puanının toplamına karşılık gelir. Güvenirlik katsayısı gerçek puan varyansının

gözlenen puan varyansına oranıdır. Klasik test kuramı paralel ölçümler yapıldığını yani madde varyanslarının eşit olduğunu varsayar (Baykul, 2000). Paralel ölçüm varsayımının sonucu olarak klasik test kuramı öncelikle bireysel farklılıkların bir teorisidir ve bireylerin göreceli durumları ile ilgilidir. Bu bireysel farklılıklarla ilgilenen güvenilirlik katsayısı, sadece bireyler ve artıklar için varyans kaynakları ile hesaplanır. Gerçek puan (T), gözlenen puan (X) olmak üzere güvenilirlik katsayısı aşağıdaki eşitlik ile gösterilir.

$$\rho (X,X')=\frac{\sigma^2(T)}{\sigma^2(T)+\sigma^2(E)} \quad (5)$$

Tek yüzeyli ve rastgele bir desen için güvenilirlik katsayısı genellenebilirlik katsayısı ile benzerlik gösterir (Atılğan, 2019; Shavelson ve Webb, 1991).

Genelleme evrenindeki ölçüm prosedürünün her birinde bireyin ortalama puanının beklenen değeri bireyin evren puanıdır. Evren puanlarının popülasyondaki tüm bireyler üzerindeki varyansına evren puan varyansı denir. Klasik test kuramında gerçek puan varyansı ile kavramsal benzerlikleri vardır. Buradan hareketle gerçek puan varyansı G kuramında birey varyansı ile yer değiştirmiş olur. G kuramında hatalar bağıl ve mutlak hata varyansları olarak farklılaştığı için iki ayrı güvenilirlik katsayısı hesaplanır. Bu katsayılardan biri "genellenebilirlik katsayısı" olarak adlandırılır ve $E\rho^2$ olarak gösterilir. Diğeri ise ϕ ile gösterilen "güvenilirlik indeksi" dir.

Genellenebilirlik katsayısı ($E\rho^2$) evren puan varyansının kendisi ve bağıl hata varyansı toplamına oranıdır ve Eşitlik 6'daki gibi gösterilir.

$$E\rho^2 = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2(\delta)} \quad (6)$$

İki yüzeyli çaprazlanmış $b \times m \times p$ desen için genellenebilirlik katsayısı, bağıl hata varyansının bireylerin diğeri yüzeyler ile ortak etki varyanslarından oluşması sebebiyle aşağıdaki gibi gösterilir.

$$E\rho^2 = \frac{\sigma^2_b}{\sigma^2_b + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{bmp,e}}{n_m n_p}} \quad (7)$$

Güvenilirlik indeksi (ϕ) evren puanı varyansının kendisi ve mutlak hata varyansı toplamına oranıdır ve Eşitlik 8’deki gibi gösterilir:

$$\phi = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2(\Delta)} \quad (8)$$

İki yüzeyli çaprazlanmış $b \times m \times p$ desen için mutlak hata varyansı birey hariç tüm yüzeylerin ortak ve ana etki varyanslarının toplamından oluşur. Bu sebeple güvenilirlik indeksi, aşağıdaki gibi gösterilir:

$$\phi = \frac{\sigma^2_b}{\sigma^2_b + \frac{\sigma^2_m}{n_m} + \frac{\sigma^2_p}{n_p} + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{mp}}{n_m n_p} + \frac{\sigma^2_{bmp,e}}{n_m n_p}} \quad (9)$$

(Atılğan, 2019; Brennan, 2001).

Güvenirlik, benzer koşullar altında tekrarlı ölçümlerde aynı sonuçların elde edilmesidir. Fakat eğitim ve psikoloji gibi sosyal bilimlerde aynı bireyler üzerinde aynı araçla aynı ölçme sonuçları elde etmek sonuçlara karışan tesadüfi hata sebebiyle pek mümkün olmamaktadır. G kuramı bir testten elde edilen ölçme sonucunu elde edilebilecek tüm sonuçlara genelleyerek bu sınırlılığı ortadan kaldırır. Bunun yanında güvenilirliği etkileyen bir diğer etken ise toplanan bilgi miktarıdır. Bir testle ilgili toplanan bilgi miktarı azaldıkça ölçme sonucu elde edilen puanın güvenilirliği dolayısıyla geçerliği azalır. Başka bir deyişle; hata kaynakları ve evrene genellenebilirliğin yanında veri miktarı da güvenilirlik üzerine çalışma yapılırken dikkat edilmesi gereken noktalardandır.

Ölçme araçlarında veriler her zaman eksiksiz toplanamamaktadır. Araştırmacılar çalışmalarının bir noktasında kayıp nicel veri sorunuyla karşı karşıya kalabilirler. Araştırmaya katılanlar soruyu cevaplamayı reddedebilir, unutabilir, verilerin kaydedilmesinde sorun yaşanabilir. Katılımcıların soruyu anlamaması, ciddiye almamaları, yorgun olmaları, sağlık sorunları gibi yanıtlayıcı kaynaklı

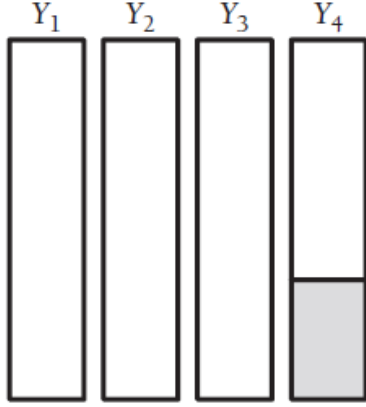
durumlar kayıp veriye sebep olabilirken; ekipman arızaları, veri girişi hataları ya da veri toplama sorunları gibi yanıtlayıcının dışındaki durumlar da eksik veri oluşmasında rol oynar. Araştırmacılar kayıp verilerle karşı karşıya kaldıklarında dikkatli davranmalıdır. Kayıp verileri analiz etme yöntemleri, verilerin niteliği ve çoğu zaman göz ardı edilen kayıp gözlemlerin nedenleri hakkında varsayımlar gerektirir. Araştırmacılar, kayıp veri yöntemlerini, bu yöntemin gerektirdiği varsayımları dikkatlice düşünmeden kullandıklarında, yanlı ve yanıltıcı sonuçlar elde etme riskiyle karşı karşıya kalırlar. Veri toplama, veri hazırlama, veri analizi ve sonuçların yorumlanması aşamalarının gözden geçirilmesi, araştırmacıların işlerinde kayıp verilerin nasıl ele alınacağı hakkında karar verirken göz önünde bulundurmaları gereken hususları vurgulayacaktır. Değişkenlerin verilerdeki dağılımı ve kayıp olma nedenleri, uygun kayıp veri tekniklerini uygulamak için iki kritik konudur (Alpar, 2003; Pigott, 2001; Tabachnick ve Fidell, 2001). Veri setlerini analiz etmek için standart istatistiksel yöntemler geliştirilmiştir. Geleneksel istatistiksel yöntemler ve yazılımlarda kullanılan veri setleri, satırları gözlemler olarak ifade edilen birimlerden, sütunları ise her birim için ölçülen değişkenlerden oluşan matrislerdir. Matristeki girdiler; yaş, gelir düzeyi gibi sürekli değişkenlerden, eğitim seviyesi gibi sıralı ya da cinsiyet, ırk gibi sıralı olmayan yanıt kategorilerinden oluşan değerlerdir. Matristeki bazı girdiler gözlenemediğinde hücre boş kalır ve bu analize olumsuz yönde etki edebilecek bir durumdur (Little ve Rubin, 2002). Geleneksel istatistiksel yöntemler modeldeki tüm değişkenlerin tüm durumlar için ölçüldüğünü varsaymaktadır. Eksik veriyi dikkate almadan yapılan istatistiksel analizlerin yansız ve güvenilir olmadığı bilinen bir gerçektir. Başka bir ifadeyle örneklemin evrene genellenebilirlik düzeyi düşüktür. Eksik veri, matris yapısını bozduğu için hatalı sonuçlar ortaya çıkar ve istatistiksel analizin kalitesini olumsuz yönde etkiler bu da ölçme puanlarının güvenilirliğini ve geçerliliğini düşürür.

Tabachnick ve Fidell'e (2001) göre kayıp verinin örüntüsü, miktarından daha önemlidir. Veri içerisinde yansız dağılan kayıp değerler, daha az miktarda fakat yansız dağılmayan değerlere göre analiz sonuçlarının genellenebilirliği açısından daha az ciddi sorun teşkil etmektedir. Başlangıç olarak kayıp veri örüntüleri ile kayıp veri mekanizmaları arasındaki farkı ayırt edebilmek önemlidir. Bu iki ifade araştırmacılar tarafından çoğu kez birbirinin yerine kullanılmaktadır. Kayıp veri

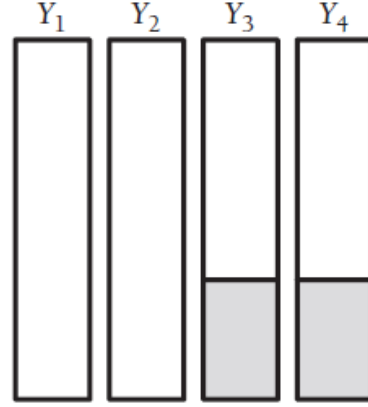
örüntüleri, kayıp değerlerin verideki konumunu gösterirken; kayıp veri mekanizmaları, söz konusu değişkenler ile kayıp değer oluşma olasılığı arasındaki ilişkiyi ifade etmektedir. Kayıp veri örüntüleri verideki 'boşlukların' sebebi hakkında bilgi vermemektedir. Kayıp veri mekanizmaları da gözlenen değer ile kayıp değer arasında nedensellik sunmasa da bunlar arasındaki matematiksel ilişkiyi ortaya koymaktadır.

Şekil 1 kayıp veri örüntülerinin eksik değerlerin verideki konumuna bağlı olarak sınıflandırılmasını göstermektedir. Kayıp değerlerin veri setindeki değişkenlerden yalnızca birinde konumlandığı tek değişken örüntüsü (univariate pattern) daha çok deneysel çalışmalarda görülür. Bazı katılımcıların cevap vermeyi reddettiği değişkenleri içeren birim yanıtlanmama örüntüsü (unit nonresponse pattern), genellikle anket çalışmalarında karşımıza çıkar. Katılımcıların deneyden ayrıldığı ve bir daha geri dönmediği çalışmalarda kayıp veriler monoton örüntü (monotone pattern) oluşturur. İlaç firmasının yaptığı bir deneyde katılımcının ilaca gösterdiği reaksiyondan dolayı çalışmadan ayrılması örnek olarak verilebilirken, bu örüntü daha çok boylamsal çalışmalarda görülür. Kayıp değerlerin yapılması belki de en yaygın olarak genel örüntüde (general pattern) görüldüğü şekildedir. Kayıp değerler veri boyunca düzensiz dağıldığı için ilk bakışta rastgele bir dağılım varmış gibi görünse de değişkenler arasında sistematik bir ilişki olabilir. Kayıp değer örüntülerinin nedensellik ilişkisi sunmadığı hatırlanarak bu tür bir belirleme kayıp veri mekanizması incelenerek bulunabilir. Veri toplama araçlarının farklı formlara ayrılıp katılımcılara uygulanmasıyla sistematik olarak kayıp veri setinin oluşturulduğu 'üç formlu anket (three-form questionnaire)' gibi yöntemlerde örüntü, planlı (planned pattern) olarak isimlendirilebilir. Planlı örüntünün oluşturulduğu veri setleri, veri toplama araçlarında çok sayıda madde olduğunda katılımcı yükünü hafifletme açısından kullanışlıdır. Son olarak gizil değişken örüntüsü (latent variable pattern), yapısal eşitlik modelleri gibi gizil değişken içeren çalışmalarda oluşur. Bu örüntüde gizil değişken, örneklemin tümü için kayıp değerdir, bu değişkene ait gözlem bulunmamaktadır (Enders, 2010).

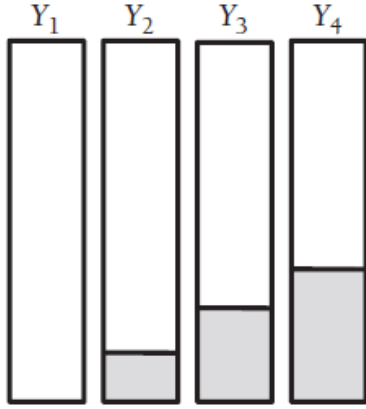
(A) Tek Değişken Örüntüsü
(UnivariatePattern)



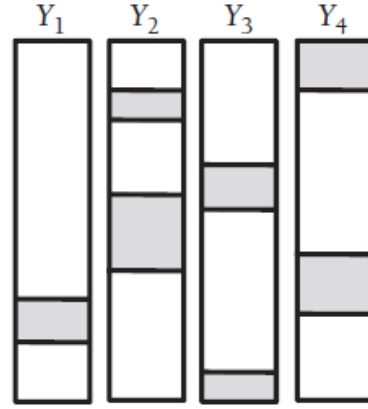
(B) Birim Yanıtlanmama Örüntüsü
(UnitNonresponsePattern)



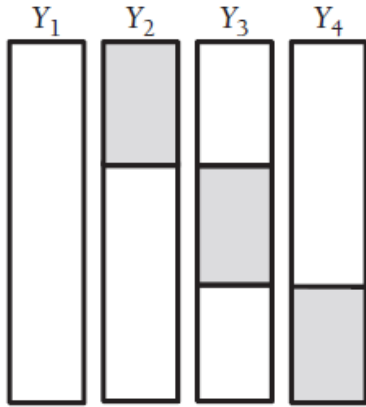
(C) Monoton Örüntü
(MonotonePattern)



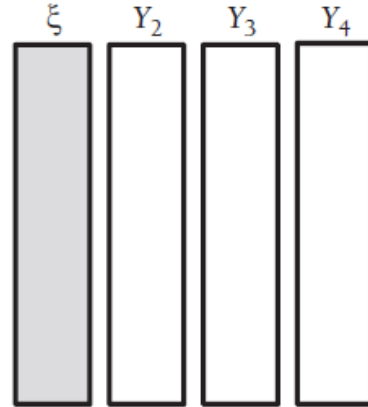
(D) Genel Örüntü
(General Pattern)



(E) Planlanmış Kayıp Örüntüsü
(PlannedMissingPattern)



(F) Gizil Değişken Örüntüsü
(LatentVariablePattern)



Kaynak: Enders, C. K. (2010). Appliedmissingdataanalysis, Newyork: TheGuilfordPress

Şekil 1. Kayıp Veri Örüntüleri

Kayıp veri mekanizmaları, değişkenlerin kayıp olma olasılığının, değişkenin kendisine mi bağlı olduğu sorusuna yanıt arar. Kayıp veri mekanizmalarının en uygun kayıp veri yönteminin belirlenmesi ve analizin doğru sonuçlar vermesi açısından çok önemlidir. Kayıp değerlere sahip verilerin analizinde mekanizmaların belirlenmesinin kritik rolü Rubin (1976) tarafından kavramsallaştırılıncaya kadar büyük ölçüde göz ardı edilmiştir. Little ve Rubin (2002) kayıp veri mekanizmalarını; Tamamen Rassal Olarak Kayıp (Missing Completely at Random, MCAR, TROK), Rassal Olarak Kayıp (Missing at Random, MAR, ROK) ve İhmal Edilemez Kayıp (Noignorable, NI, İEK) olarak kategorize etmişlerdir.

$Y = (y_{ij})$ tam veri matrisi, $M = (m_{ij})$ kayıp veri gösterge matrisi ve ϕ kayıp veri parametresi olarak tanımlanırsa; kayıp veri mekanizması verilen Y nin M üzerine koşullu dağılımı ile karakterize edilir ve $f(M|Y, \phi)$ ile gösterilir. Kayıp olma durumu gözlenen veya eksik tüm Y, ϕ için $f(M|Y, \phi) = f(M|\phi)$ eşitliğini sağlıyorsa veri TROK olarak adlandırılır.

Daha açık bir ifadeyle; herhangi bir değişkende eksik değer olma olasılığı, ilgili modeldeki diğer değişkenlere veya değişkenin diğer değerlere bağlı değildir (Allison, 2001; Little ve Rubin, 2002). Gözlenen değerler, verilerin eksiksiz olması durumunda analiz edilebilecek puanların rastgele bir örneğidir ve gözlenen veride yanlılığa sebep olmaz (Alpar, 2003; Enders, 2010). Y_{obs} , Y 'nin gözlenen birimlerini veya girdilerini, Y_{mis} ise kayıp bileşenlerini gösterdiğinde TROK'tan daha zayıf bir varsayım olan ROK, değişkendeki eksikliğin Y 'nin gözlenen bileşenlerine (Y_{obs}) bağlı olduğunu, eksik olan bileşenlerine bağlı olmadığını gösterir. Formel olarak ifade etmek gerekirse; tüm Y_{mis}, ϕ için $f(M|Y, \phi) = f(M|Y_{obs}, \phi)$ 'dir (Little ve Rubin, 2002).

ROK, TROK mekanizmasının özel bir halidir. Yani TROK aynı zamanda ROK varsayımını da sağlar. Analizdeki diğer değişkenler kontrol altına alındığında bir değişkenin eksik olma olasılığı, eksik veri içeren değişkenin gözlenmemiş değerleri hariç gözlemlenen herhangi bir değere bağlı olabilir (Allison, 2001). Bununla birlikte ROK varsayımı bir veya daha fazla ölçülen değişken ile eksik verilerin olasılığı arasında sistematik bir ilişki olduğu anlamına gelir. Örneğin; bir psikoloğun bir grup kanser hastasında yaşam kalitesini araştırdığını ve yaşlı hastalar

ile eğitim seviyesi düşük hastaların yaşam kalitesi anketini reddetme eğiliminin daha yüksek olduğunu tespit edilmiş olsun. Eksik veri eğilimi ve değişkendeki eksiklik arasında kalıcı ilişki olmadığı sürece veri ROK olarak nitelendirilebilir (Enders, 2010). Yaş ve gelir düzeyi değişkenlerinin olduğu bir veri setinde gelir düzeyi değişkeninin eksik değer içerdiğini, yaş değişkeninin tam olduğunu düşünelim. Gelir düzeyi değişkenindeki kayıp değerlerin olasılığı, tüm yaş ve gelir düzeyleri için eşit ise verinin TROK, gelir düzeyinde kayıp değer olma olasılığı yaşa bağlı olarak düşük veya yüksek çıkabilir fakat gelir düzeyi değişkeni ile bağlantılı değil ise verinin raslantısal olduğu söylenebilir (Alpar, 2003) .

TROK veya ROK olmayan kayıp veri raslantısal olmayan kayıp olarak adlandırılır. Eksik veri olasılığı, diğer değişkenleri kontrol ettikten sonra bile değişkenin değeriyle ilgili olduğunda veriler rastgele eksik değildir. Bu durumda eksik veriler göz ardı edilemez ve geçerli sonuçlar elde edebilmek için kayıp veri mekanizmasının modellenmesi gerekir (Allison, 2001; Enders, 2010). Önceki örnek üzerinden gidilirse, gelir düzeyi değişkeninde kayıp değer olma olasılığı hem yaş hem de değişkenin kendisinden etkilenir ise ver İEK olarak adlandırılır. Belirli yaş grubundaki bireyler gelir düzeyini belirtme konusunda isteksiz davranabileceği gibi, gelir düzeyi yüksek olan bireyler de soruyu yanıtlama konusunda isteksiz davranabilirler.

ROK varsayımı test edilebilir değildir. Eksikliğin olasılığının eksik olan değerlerle ilişkili olduğu düşünülebilir. Fakat bu ilişkiyi inceleyecek herhangi bir veri bulunmamaktadır. ROK varsayımının test edilebilmesi için kayıp değerlerin gerçek değerlerinin bilinmesi gerekmektedir (Allison, 2001). ROK mekanizması gibi kayıp değerleri bilmeden İEK mekanizmasını doğrulamak mümkün değildir (Enders, 2010).

ROK ve İEK varsayımlarının aksine TROK varsayımı test edilebilirdir. Verideki bir değişkene ait gözlemler, kayıp veri bulunanlar ve bulunmayanlar olarak iki gruba ayrılır. Gösterge değişken yardımıyla bu iki grup arasında anlamlı bir fark olup olmadığı araştırılır. Bu istatistiksel inceleme t testi kullanılarak yapılır. Gözlem yapılamayan birimler ile gözlem yapılan birimler arasında anlamlı bir farkın bulunması rastgele olmayan eksik sürece işaret eder. Anlamlı bir fark bulunmaması herhangi bir sistematik farkın bulunmadığını gösterir. Rastgeleliği belirgin hale

getirmek için diğerk deęişkenler de gösterge deęişken olarak ele alınmalıdır. Tamamen rastgelelięi incelemek için ise deęişkende eksik veri olma olasılıęının deęişkenin kendisi ile iliřkili olmadıęının kanıtlanması gerekir (Alpar, 2003).

Ayrıca TROK varsayımını test etmenin bir diğerk yolu Little's MCAR testidir. Bu test her deęişkendeki ortalama farkları eř zamanlı olarak deęerlendiren çok deęişkenli bir t testi yaklařımıdır. Tek deęişkenli t testlerinin aksine bu tüm veri kümesine uygulanan bir TROK testidir. Little's MCAR testi $p < .005$ olması durumunda verinin TROK olmadıęını gösterir (Enders, 2010).

Kayıp verilerin rastgelelięinin arařtırılmasından sonraki adım kayıplıęı giderecek yöntemlere karar vermektir. Davranıřsal ve demografik maddeleri içeren bir anket formunda gelir ile ilgili soruları yanıtlamayan bireylerin bu eęilimi bu deęişkene ait tavırla ilgili olabilir. Bu bireylerin bilgilerini analizden çıkarmak bu deęişken ile ilgili tavra yönelik örneklemede ciddi kayıplara sebep olabilir. Bu tür durumlarda kayıp verilerin kestirimine yönelik bazı yöntemler kullanılması gerekmektedir. Kayıp verilerin daęılımı TROK yapıda olduęunda kayıp veriler kestirilemez ve bu en avantajlı durumdur. ROK yapıda olan kayıp veriler diğerk deęişkenler yardımıyla kestirilebilir. İEK daęılıma sahip kayıplarda bu olasılık deęişkenin kendisi ile baęlantılıdır ve göz ardı edilmesi yanlı sonuçlara sebep olur. Kayıp deęerlerin veri içerisindeki daęılımı yansız ve az miktarda (%5 veya daha az) ise kayıp verilere uygulanan gerekli yöntemlerden elde edilen sonuçlar neredeyse aynı çıkar. Fakat orta ve küçük ölçekli örneklemlerde fazla sayıda kayıp veri var ise sorun daha ciddi olabilir. Toler edilebilecek kayıp veri oranı konusunda net bir kural yoktur (Tabachnick ve Fidell, 2013). Kayıp verinin nasıl yönetileceęi ile ilgili karar vermek önemlidir. Bu problemi ortadan kaldırmak için arařtırmacılar kayıp veri ile bař etme yöntemleri geliřtirmişlerdir. Bu yöntemlerden en geleneksel olanları Liste veya Durum Bazında Silme (LBS), Çiftler Bazında Silme (ÇBS) olarak isimlendirilen silmeye dayalı tekniklerdir. Kayıp Gözlem ile Tam Gözlemin Yer Deęiřtirmesi, Kayıp Gözlemin Ortalama ile Yer Deęiřtirmesi (OA), Hot Deck Atama, Cold Deck Atama, Regresyon Atama (RA), Çoklu Atama (ÇA), Beklenti-Maksimizasyon (BM) ise atamaya dayalı en sık kullanılan yöntemlerdir. İhtiyaç duyulan yönteme karar verilirken önyargı en aza indirilmelidir. Kayıp veri, parametre tahminlerinde yanlılıęa sebep olsa bile seçilen iyi bir yöntem bu yanlılıęı

mümkün olduğu kadar düşürmelidir. Verimli parametre tahminlerini üretebilmek için eldeki bilginin tamamı mümkün olduğu kadar kullanılmalı ve bilgi atmaktan kaçınılmalıdır. Standart hataların, güven aralıklarının ve p değerlerinin doğru tahminleri elde edilmeye çalışılmalıdır. İyi bir kayıp veri yöntemi kayıp veri mekanizmasının kısıtlayıcı varsayımlarına karşı gelişmiş olmalıdır. Rastgelelik konusunda her tür kayıp veri matrisine uygun olmalıdır (Allison, 2001). Yukarıda bahsedilen kayıp veri ile baş etme yöntemleri ile bu yöntemlerin avantajları ve dezavantajları aşağıda açıklanmıştır.

Liste veya Durum Bazında Silme

Bu yöntemde bir veya daha fazla sayıda eksik olan durumlar silinir, sadece tam olan gözlemler kullanılır. Bu sebeple "tam durum analizi" olarak da bilinir. Bu yöntemin en büyük avantajı kolaylığıdır. Analizleri tam verilerle yapmak karmaşık yazılımlara ve kayıp veri sürecine olan bağımlılığı ortadan kaldırmaktadır. Ayrıca tüm analizler için ortak bir veri seti elde edilmiş olur. Ancak çoğu durumda dezavantajları avantajlarına ağır basmaktadır. Bu yöntemin temel sorunu, TROK yapıda kayıp veri gerektirmesi ve olmadığı durumlarda yanlış parametreler üretmesidir (Enders, 2010). Yani bu yöntem sadece TROK yapıdaki verilerde kullanılmalıdır aksi durumlarda veride rastgele olmayan kayıp değerler bulunmaktadır. Çünkü TROK mekanizması, mevcut olan tam gözlemlerin orijinal gözlemlerin rasgele bir örnekleme olduğunu kabul etmektedir ve kayıp değerlerin sonuçları yanlışlaştırmayacağını sadece güven aralıklarında artışa sebep olacağını varsaymaktadır. TROK olmayan yapıdaki eksik değerler analizden çıkarıldığında bile yanlış parametrelere neden olan rastgele olmayan eksik değerler yüzünden sonuçlar evrene genellenemez (Alpar, 2003). Veriler ROK yapıda ise bu yöntem yanlış sonuçlar üretmektedir. Örneğin; kadınların % 85 inin erkeklerin ise sadece % 60'ının yanıtladığı gelir düzeyi ile ilgili bir soruda, her cinsiyet için gelir değişkenindeki eksikliğin gelire bağlı olmadığı ROK yapıda bir veri seti düşünelim. Erkeklerin kadınlardan daha fazla kazandığı düşünülürse, liste bazında silme yöntemi tüm popülasyon için ortalama gelirin aşağı yönlü bir tahminini üretir (Allison, 2001). LBS, özellikle atılan durumlar çok sayıda değişken içerdiğinde veride çok fazla

kayba neden olur. Eksik veri kayıtlarının silinmesi ile toplam veri büyüklüğünde; eksik veri oranı veya değişken sayısı arttıkça büyüklüğü artan bir azalma görülebilir. Doğaldır ki örneklem büyüklüğünün azaltılması, küçük ve orta düzeydeki veri setlerinde istatistiksel gücü önemli ölçüde azaltır. TROK yapıdaki verilerde bile bu sorundur (Enders, 2010). TROK yapıdaki veriler için tam durum analizi, tek bir ölçme ile evrene genelleştirilebilen sonuçlar sağlar. Fakat tahminler, araştırmacılar tarafından başlangıçta planlanandan daha az kesin olacaktır, çünkü tahmin için kullanılan örneklem sayısında azalma olmuştur (Pigott, 2001).

Çiftler Bazında Silme

Çiftler bazında silme, eldeki tüm verinin kullanılmasına dayanan LBS yöntemine alternatif bir yöntemdir. Bu yöntem eksik olan verileri silme ya da yerlerine değer atamaya değil, eldeki tüm değerler yardımıyla tanımlayıcı ve ilişkisel ölçülerin elde edilmesini sağlar. Daha çok doğrusal modellerde (doğrusal regresyon, faktör analizi) kullanılan yöntemde ilgilenilen parametreler ortalama, korelasyon ve kovaryanstır. Buradan elde edilen katsayılar tüm örnekleme genellenir. Çiftler bazında silmede her değişken veya değişken çifti için mevcut tüm veriler kullanılarak tahmin yapılır. Bu şekilde verilerin hepsi kullanılmış olur ve hiç bir veri kaybı yaşanmaz. Örneklemedeki gözlem sayısı maksimize edilerek korelasyon katsayıları elde edilir. Bu yöntemin en belirgin özelliği; her bir gözlem çifti kadar korelasyon katsayısı elde edilmesi ve her birindeki mevcut gözlem sayısının farklı sayıda olmasıdır. Yöntemdeki başlıca sorun; korelasyonların beklenen sınırlarından farklı çıkması ve korelasyon matrisindeki diğer korelasyonlarla tutarsız olmasıdır. Diğer bir sorun ise korelasyon matrisinin özdeğerlerinin 0'dan küçük çıkabilmesidir bu da matrisin varyansını etkilemektedir.

LBS gibi ÇBS de veri TROK yapıda değilse yanlış sonuçlar üretmektedir. Eldeki tüm verinin kullanılması açısından LBS yöntemine göre daha verimli sonuçlar elde edileceği düşünülse de bazı simülasyon sonuçları ÇBS nin daha az etkili olabileceğini göstermiştir (Allison, 2001; Alpar, 2003).

Kayıp Gözlem ile Tam Gözlemin Yer Değiřtirmesi

Bu yöntem en genel tanımıyla örnekleme de yer alan hiç bir maddeyi yanıtlamamış veya çok fazla eksik bilgisi bulunan kişilerin kayıp deęerlerinin yerine, örnekleme giren ve bu kişiler ile özellikleri benzer olan kişilerin deęerlerinin atanmasıdır. Yedek denek atama teknięi de denmektedir (Alpar, 2003).

Kayıp Gözlemin Ortalama ile Yer Deęiřtirmesi

Arařtırmacılar tarafından sıklıkla kullanılan kayıp gözlemlerin ortalama ile yer deęiřtirmesi dięer bir deyişle ortalama atama; popüler bir yöntem olmasına rağmen güncel yazılım programlarının sunduęu kolaylıklar sebebiyle daha az tercih edilir hale gelmiştir. Bu işlem; incelenen deęiřkene ait tam gözlemlerin ortalaması alınarak (seri ortalaması) dikey bir şekilde, duruma ait gözlemlerin ortalaması alınarak (gözlem birimleri ortalaması) yatay bir şekilde yapılabilir. Burada bahsedilen seri ortalaması atama yöntemidir. Bu yöntemde ilgilenilen deęiřkene ait tam gözlemlerin ortalaması o deęiřkene ait kayıp veriler ile yer deęiřtirir. Kayıp deęerlerle ilgili elde hiç bir bilgi yoksa en iyi tahmin kayıp deęerlerin ortalamasıdır. Yöntemin tercih ediliyor olma sebeplerinden biri, kayıp deęerler yerine atanan deęerlerin daęılımının ortalamasını deęiřtirmiyor oluşu ve arařtırmacıya tahmin zorunluluęu vermemesidir. Bununla birlikte, kolaylık ilgi uyandıran bir avantaj deęildir çünkü bu yöntem, veriler TROK olsa bile sonuçta ortaya çıkan parametre tahminlerini ciddi şekilde bozmaktadır.

Deęiřkendeki kayıp veriler ortalama ile yer deęiřtirdięinden yanlı parametreler üretilir, gerçek deęerden uzaklaşılır ve deęiřkenin varyansı düşmektedir. Bu durum varyansa baęlı olan regresyon katsayıları vb. gibi başka parametreleri de etkilemektedir. Varyans kaybının büyüklüęü veri kayıp miktarına ve kayıp verinin gerçek deęerine baęlıdır. Verilerin deęiřkenlięinin sınırlandırılması aynı zamanda kovaryansların ve korelasyonların büyüklüęünü de azaltmaktadır (Allison, 2001; Alpar, 2003; Enders, 2010). Kayıp deęerlerin ortalama gibi tek bir deęerle deęiřtirilmesi, muhtemel varyansı azaltarak bu deęiřkenin daęılımını deęiřtirir. Varyansın azalmasının sebebi; kayıp deęerlerin aynı ortalama deęerle

doldurulmasının, deęişkenler gözlenirse muhtemel olabilecek varyansları hesaba katmamasıdır. Gerçek deęerler muhtemelen ortalamadan farklılık gösterecektir. Dięer bir sebep ise artan örneklem büyüklüğünden kaynaklanan daha küçük standart hatalar, verilerde mevcut belirsizlięi yeterince yansıtmamasıdır (Pigott, 2001).

Kayıp deęerlerin grup ortalaması ile deęiştirilmesi bir dięer yoldur. Bu yol; kayıp veriyi deęişkendeki tüm verinin ortalaması ile deęiştirmeye göre biraz daha risklidir fakat tahmin daha güvenilirdir. Ancak yöntem; grup içi farklılıkların düşmesine, gruplar arası farklılıkların artmasına neden olmaktadır (Tabachnick ve Fidell, 2013).

Geleneksel atama yöntemlerinin genel olarak standart hatayı düşürme eğiliminde olması abartılı istatistik deęerlere sebep olmaktadır. O yüzden bu yöntemlerin 'dürüst olmayan (nonhonest)' yöntemler olduęu belirtilmektedir (Allison, 2001).

Hot Deck Atama

Bu yöntemde 'benzer' yanıt verenlerin deęerleri eksik deęere atanır. Veri setindeki dięer gözlemlerden yararlanılır. Eksik verinin bulunduęu birimdeki tam olan gözlemlere benzeyen başka birimler arasından rastgele biri seçilir. Bu birimdeki ilgili deęişkene ait deęer eksik verideki eksik deęere atanır. Davranışsal ve sosyal bilimlerde nadiren kullanılan yöntem, anket literatüründe oldukça fazla dikkat çekmiştir ve uzun bir tarihe sahiptir.

Hot Deck atama yönteminin uygulamasında çeşitli yollar mevcuttur. En basit olarak, gözlemlenen verilerden rastgele alınan ilgili deęişkene ait deęer(ler) kayıp deęer(ler) yerine atanır. Bu yöntemin daha tipik bir uygulamasında eksik deęer, bazı eşleştirilen deęişkenlerde benzer yanıt verenlerin alt örneęi ile deęiştirilir. Örneęin bazı katılımcıların gelir düzeyini yanıtlamaktan kaçındığı bir araştırmada Hot Deck prosedürü katılımcıları yaş, cinsiyet, ırk ve medeni durum gibi demografik özelliklere göre sınıflandırır. Daha sonra eksik deęer, eksik verileri olan bireyle aynı demografik özelliklere sahip katılımcıların gelir düzeyi yanıtlarından rastgele bir deęer ile deęiştirilir.

Fazla sayıda benzer eksik deęerin olduęu durumlarda, benzer gözlemlerin oluşturduęu kümenin büyük olması aynı gözlemlerin tekrar tekrar kullanılmasının önüne geçer. Eksik deęer içeren gözleme benzeyen gözlemlerin oluşturduęu kümenin küçük olması da 'benzerlik' kavramının tam olarak kavranabilmesi açısından olumludur. Yöntemin olumlu yanı kavramsal olarak kolay anlaşılıyor olması ve tamamlanan deęişkenin varyansını azaltmamasıdır. Bunun yanında yöntem yanlı korelasyon ve regresyon katsayı tahminleri üretebilmektedir. Dięer atama yöntemlerinde olduęu gibi standart hatayı hafife alma eğilimindedir (Alpar, 2003; Enders, 2010; Little ve Rubin, 2002).

Cold Deck Atama

Bu yöntemde; aynı anketin daha önceki uygulamasından elde edilen bir deęer gibi harici bir kaynaktan gelen deęer, kayıp deęerin yerine atanır. Yöntemin kayıp deęer ile ortalamanın yer deęiştirdięi yöntemden tek farkı atanan deęerin başka bir uygulamadan elde edilmesidir. Yer deęiştirme yöntemlerinde olduęu gibi, mevcut uygulama genellikle sonuç olarak elde edilen verileri eksiksiz bir örnek olarak ele alır, yani atama sonucunda ortaya çıkan yanlılığı dikkate almaz. Cold Deck atama, kayıp deęer ile ortalamanın yer deęiştirmesi yöntemi ile benzer olumsuzluklara sahiptir (Alpar, 2003; Little ve Rubin, 2002)

Regresyon Atama

Bu yöntemde veri setindeki kayıp deęerler, gözlenen ve eksik deęişkenlerin bulunduęu birimlerdeki gözlenen deęişkenler yardımıyla tahmin edilen deęerler ile tamamlanmaktadır (Little & Rubin, 2002). RA, kayıp deęerlerin bağımlı deęişken, dięer deęişkenlerin bağımsız deęişken olarak kullanıldıęı regresyon çözümlmesine dayalı bir yöntemdir. Veri setindeki gözlenen deęerler regresyon denklemini oluşturur ve bu denklemden elde edilen deęerlerden kayıp deęerler kestirilir. RA, araştırmacının bilinçsizce yaptıęı tahminin önüne geçmesi bakımından daha objektif bir yöntemdir (Tabachnick ve Fidell, 2013).

Bu yaklaşımın arkasındaki temel fikir eksik değişkenleri doldurmak için tüm değişkenlerden bilgi kullanıyor olması bakımından sezgisel olarak ilgi çekicidir. Değişkenler arasında korelasyon kurulma eğilimindedir, bu nedenle gözlemlenen verilerden bilgi alan atamalar üretmek mantıklıdır (Enders, 2010).

Yöntem, eksik veri orta düzeyde ve veri içine yaygın olarak dağılmış ise kullanılmalıdır. Bunun yanında değişkenler arası ilişki yeterli düzeyde olmalı, bağımlı değişken bağımsız değişken tarafından iyi düzeyde açıklanabilmelidir. Aksi halde OA yönteminin kullanılması daha verimlidir (Alpar, 2003). RA yönteminin olumlu yanlarının yanında olumsuz tarafları da bulunmaktadır. Bunlar şöyle sıralanabilir:

- Eldeki değerlerden kestirim yapılacağı için var olan ilişki daha da kuvvetlenmiş olur bu da gerçek hayatta olan değerlerden daha yakın değerler üretir.
- Kayıp değerler eldeki değerlere benzeyeceği için varyans azalmaktadır.
- Kayıp değerlerin kestiriminde kullanılan bağımsız değişkenlerin kaliteli olması da önemlidir. Bağımsız değişkenler kayıp verileri doğru bir şekilde kestiremezse RA'nın OA'dan farkı bulunmamaktadır.
- Regresyon sonucu elde edilen kestirimin sınırları var olan sınırların ötesine geçebilmektedir (Tabachnick ve Fidell, 2013).

Beklenti - Maksimizasyon

Beklenti maksimizasyon, kestirilen değerlerdeki değişimin önemsenmeyecek derecede azalmasına kadar devam eden beklenti (B) ve maksimizasyon (M) aşamalarından oluşan yinelemeli (iteratif) bir yöntemdir. RA yöntemi gibi araştırmacılar tarafından sıklıkla kullanılmaktadır. Beklenti basamağında, beklentiye hesaplamak için parametre kestirimlerinin eldeki değerleri kullanılarak kayıp değer içeren değişkenler üzerinden beklenen logaritmik olabilirlik değerleri elde edilir. Maksimizasyon basamağında ise yeni parametre kestirimleri elde etmek için beklenen logaritmik olabilirlik değerleri maksimize edilir. Bu iki adım elde edilen

parametre kestirimleri bir yinelemeden diğere deđiřmeye kadar tekrarlanır. BM algoritmasında çok deđiřkenli normal dađılımlı bir modelde kestirilen parametreler; ortalama, varyans ve kovaryanstır. BM yönteminde adımlar genel olarak ařađıdaki gibidir:

1. Ortalama, varyans ve kovaryans eldeki kayıp olmayan veriler yardımıyla hesaplanır. Bunlar liste bazında silme veya çiftler bazında silme yöntemleri ile elde edilebilir.
2. Her kayıp veri seti için, kayıp verileri tahmin etmek için gözlenen deđiřkenler kullanılarak regresyon denklemleri oluşturulur. Regresyon parametreleri bir önceki adımda elde edilen ortalama, varyans ve kovaryansların mevcut tahminlerinden doğrudan hesaplanır.
3. Bu regresyon denklemleri, tüm deđiřkenler ve eksik veri içeren durumlara deđer üretmek için kullanılır.
4. Tüm gerçek ve tahmin verileri kullanılarak ortalama, varyans ve kovaryans parametreleri yeniden hesaplanır.
5. 2. aşamaya geri dönülür ve kestirilen deđerler yakınsayana kadar algoritma tekrarlanır.

Kestirilen parametre deđerleri tahmin sürecinin bir parçası için kullanılmaktadır. Başka parametrelerin sürece dahil edilmesi yanlı sonuçlar doğurabilmektedir. Doğru parametre tahminleri üretmesinin yanında BM algoritmasının olumsuz yanı, standart hata kestirimi üretmemesidir (Allison, 2001; Alpar, 2003).

Çoklu Atama

En az iki atama yönteminin birlikte kullanıldığı çoklu atama yönteminde kayıp veri, karma bir kestirim deđerı olarak tahmin edilir. Karma kestirim deđerı birden fazla atama yöntemiyle elde edilen kestirim deđerlerinin ortalamasıdır. BM yönteminde olduđu gibi bu yöntemde de kayıp veriler regresyon denklemleri yardımıyla kestirilir. Fakat bu atamada basit atama yöntemlerinde olduđu gibi tek veri deđer $m \geq 2$ olacak şekilde m tane tamamlanmış veri seti üretilmektedir. Daha

sonra m tamamlanmış veri seti üzerinde standart istatistiksel analizler ayrı ayrı uygulanmaktadır. Farklı sonuçlardan elde edilen parametre kestirimlerinin ortalaması alınarak yorum yapılır (Alpar, 2003; Little ve Rubin, 2002; Tabachnick ve Fidell, 2013).

Hemen hemen her veriye ve modele uyumlu olması ve geleneksel yazılımlarla uygulanabilir oluşu ÇA yönteminin avantajları arasındadır. Dezavantajları arasında en belirgin olanı ise her seferinde farklı sonuçlar üretiliyor olmasıdır. İkinci bir dezavantaj; belirsizliğe ve karışıklığa yol açacak şekilde atama yöntemlerinin farklı yollarının olmasıdır. Çeşitli yaklaşımlar arasında karar vermek araştırmacılar için zor olabilmektedir. En çok kullanılan yöntem doğrusal regresyona dayanan Markov Zinciri Monte Carlo (MCMC) algoritmasıdır. MCMC algoritması BM algoritmasına benzer şekilde işler. Tek farkı ise kestirilen değerlere eklenen rastgele varyanslardır. Bu, atama yöntemlerinde sıklıkla karşılaşılan varyans tahminlerindeki aşağı yönlü yanlılığı telafi eder. Kayıp veri tek bir değişkende ise kullanışlı olan yöntem parametre ve standart hataların doğru tahminlerini üretmek yönünden başarılıdır. Çok değişkene dağılmış kayıp veride yöntemi kullanmak güçleşmektedir (Allison, 2001).

Değişkenler arasındaki istatistiksel ilişki, kayıp değişkenlerde kaybedilen bilgilerin tamamen gözlemlenen diğer değişkenler tarafından desteklenmesi olasılığını arttırmaktadır. Model tabanlı olarak nitelendirilen çoklu atama ve beklenti maksimizasyon yöntemleri , kayıp verileri işlemek için değişkenler arasındaki çok değişkenli ilişkiyi kullanır. Bu nedenle, veri seti ne kadar bilgilendirici olursa, önemli yapılar üzerinde ne kadar çok önlem alınırsa model tabanlı yöntemler kullanarak yapılan tahmin o kadar iyi olacaktır.

Veri toplama sırasında araştırmacı, kayıp verilerin nedenlerini gözleme ve özellikle kayıp değerlere duyarlı değişkenler için daha fazla bilgi toplama olanağına sahiptir. Tam durum analizi ve model tabanlı yöntemler, yalnızca kayıp veri mekanizması ve verilerin dağılımına ilişkin varsayımlar kabul edildiğinde güvenilir sonuçlar sağlar (Pigott, 2001).

Yukarıda bahsedilen kayıp veri baş etme yöntemlerinden hem LBS, OA ve RA gibi geleneksel yöntemlerin hem de BM, ÇA gibi yeni yaklaşımların avantaj ve dezavantajları bulunmaktadır. Gerçek anlamda hiç birisi tam olarak iyi şekilde

nitelendirilemez. Kayıp veri için gerçek çözüm mümkün olduğu kadar az eksik gözlem içeren veri seti elde etmektir. Ancak bu yöntemlerden bazılarının diğerlerinden daha iyi olduğu söylenebilir (Allison, 2001).

Ölçme aracının psikometrik özelliklerinden güvenilirlik üzerine çalışma yapılırken kullanılan veri setinde de eksik gözlem sonucu kayıp veriler oluşabilir. Kayıp veri üzerinde bazı çalışmalar yapmak gerekmektedir. Elde edilen sonuçların yansızlığı, evrene genellenebilirliği ve kestirimlerin doğruluğu, yapılan analizlerde kayıp veri sorununun ne derece dikkate alındığına bağlıdır. Kayıp verilerin dikkate alınması ve bunun için önlem alınması analizin titizlikle yapıldığının göstergesidir. Bu durum G kuramı analizleri içinde geçerlidir. Bu nedenle G kuramında varyans bileşenlerinin ve güvenilirlik katsayılarının kestirilmesi için kullanılan veri setinde kayıp veri olması durumunda; kayıp veri ile baş etme yöntemlerinin karşılaştırmalı olarak incelenmesi gereklidir. Böylelikle G kuramı analizlerinde hangi kayıp veri ile baş etme yönteminin daha gerçekçi sonuçlar elde edilmesini sağladığına ilişkin kuramsal bilgiye ulaşılması önemlidir. Bu araştırmada, G kuramında varyans bileşenlerinin ve güvenilirlik katsayılarının kestirilmesi için kullanılan veri setinde kayıp veri olması durumunda; kayıp veri ile baş etme yöntemlerinin karşılaştırmalı olarak incelenmesi için aşağıdaki problem cümlesine ve alt problemlere yanıtlar aranmaya çalışılmıştır.

1.1 Problem Cümlesi

Aynı veri setinden farklı oranlarla oluşturulmuş eksik verili veri setlerinden farklı kayıp veri ile baş etme yöntemlerinin kullanımıyla elde edilen G katsayısı, Phi katsayısı, evren puanı varyansı, mutlak hata varyansı ve bağıl hata varyansı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

1.1.1 Alt Problemler

1. Aynı veri setinden %5 oranlarla oluşturulan eksik verili veri setlerinde;

a) “LBS” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

b) “OA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

c) “RA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

2. Aynı veri setinden %10 oranlarla oluşturulan eksik verili veri setlerinde

a) “LBS” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

b) “OA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

c) “RA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

3. Aynı veri setinden %20 oranlarla oluşturulan eksik verili veri setlerinde

a) “LBS” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

b) “OA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

c) “RA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

1.2 Araştırmanın Amacı ve Önemi

Bu çalışmada kayıp verilerle baş etmede kullanılan yöntemlerin G kuramı çok değişkenli analizi ile elde edilen sonuçlar üzerindeki etkilerinin farklı kayıp veri oranları dikkate alınarak incelenmesi amaçlanmaktadır.

Birey (b), madde (m) ve puanlayıcı (p) değişkenlik kaynaklarıdır. Çalışmada tüm bireylerin tüm maddeleri yanıtladığı, tüm maddelerde tüm puanlayıcılar tarafından puanlandığı üç değişkenlik kaynağından oluşan iki yüzeyli (facet) çaprazlanmış desen kullanılmıştır. Desen, $bxm \times p$ olarak tanımlanmıştır. Tamamen yansız %5, %10 ve %20 oranlarında kayıp değerlerin olduğu kayıp veri mekanizmasındaki veri setlerinde LBS, OA ve RA yöntemleri ele alınmıştır.

Genellenebilirlik ile güvenilirlik kestirimi çalışmalarına başlamadan önce kayıp veri sorununun çözülmesi gerekmektedir. İstatistiksel yazılımların çoğunda sıklıkla kullanılan metot eksik veri içeren değişkenleri veya durumları veriden çıkarmaktır. Örneklemin büyük bir kısmını silmek, istatistiksel olarak güç kaybına neden olmaktadır. Araştırmacılar, toplamak için çok fazla zaman, çaba ve para harcadıkları verileri atma konusunda isteksiz davranırlar. Bu nedenle eksik değerlerin veriye kazandırılması için birtakım yöntemler kullanılmalıdır (Allison, 2001).

Verideki eksik değerlerin oranı arttıkça, toplanan bilgi azalacak ve ölçme sonuçlarının güvenilirliği azalacaktır. Ölçme aracından toplanan puanların doğru bir şekilde yorumlanabilmesi için kayıp veri sorununun çözümlenmesi gerekmektedir. Kayıp veri ile ilgili çalışmalar incelendiğinde; farklı kayıp veri yöntemleri ve farklı oranlardaki kayıp değerler, elde edilen bulgularda farklılığa sebep olmaktadır. Farklı oranlarda eksik veriye bağlı olarak kayıp veri yöntemlerinin karşılaştırıldığı bu çalışmadan elde edilen bulgularla, G kuramı çok değişkenli analiz sonuçlarının en doğru şekilde elde edilmesinde veri setine uygun yöntemin seçilebilmesi amaçlanmıştır.

2. İLGİLİ ARAŞTIRMALAR

Bu bölümde, ilgili alan yazında kayıp veri ile baş etme yöntemlerinin ölçme araçlarının güvenilirliğine ve G kuramına etkisi ile ilgili Türkiye'de ve dünyada yapılan çalışmalara yer verilmiştir.

Enders (2003), LBS, ÇBS, OA, bireysel ortalama atama (BOA) ve BM yöntemlerini farklı kayıp veri koşulları altında Cronbach-alfa katsayısı temelinde incelemiştir. TROK, ROK ve İEK koşulları altındaki verilerde madde sayısı ($k=10$ ve $k=20$), cevap kategorileri ($c=3,5$ ve 7), maddeler arası korelasyonlar ($r=.301$ ve $r=.562$) örneklem büyüklüğü ($N=100, 300$ ve 500), kayıp veri oranı (%15-%30) değişkenleri manipüle edilerek Cronbach-alfa katsayıları yanlılık, ortalama hatanın karekökü ve güven aralığı temelinde incelenmiştir. Yapılan analizlerin sonucunda elde edilen mevcut sonuçlar BM yönteminin daha kesin kestirimler ürettiğini göstermiştir. Bu, elde edilen alfa katsayısının referans değere yakın olduğunu göstermiştir. Tüm kayıp veri koşullarında yöntemler etkililiğine göre sıralanırsa BM ve BOA ilk iki sırayı almıştır. TROK koşulu altında bile LBS yöntemi negatif yanlı kestirimler üretmiştir. İEK koşulu altında BOA yöntemi BM yönteminden daha iyi sonuçlar elde edilmiştir. BM kalan kayıp veri yöntemlerinden daha iyi olmasına rağmen tüm yöntemlerde elde edilen tahminler negatif yanlı bulunmuştur.

Enders (2004), 'Örnek Güvenirlik Kestirimleri Üzerinde Kayıp Verinin Etkisi: Güvenirlik Raporlama Uygulamaları Üzerine Tavsiyeler' adlı çalışmasında madde düzeyinde eksik verilerle güvenirlik analizlerine En Çok Olabilirlik (EÇO) yaklaşımını dahil etme yönteminin önemini belirtmiştir. Araştırmanın öncelikli amacı EÇO yaklaşımını kullanarak madde düzeyinde eksik verilerle güvenirlik analizi yapmak için iki aşamalı bir yöntemi ana hatlarıyla belirtmektir. Diğer bir amaç LBS yöntemi gibi geleneksel yöntemlerin güvenirlik tahminleri üzerine olumsuz etkisini simülasyon çalışması ile göstermektir. Son olarak ise güvenirlik raporlama uygulamalarının etkilerini tartışmaktır. Bu amaçla TROK ve ROK mekanizmasına sahip 1000 adet yapay veri seti oluşturulmuştur. Toplamda 2000 çoğaltmanın her biri LBS, ÇBS, BM ve OA gibi kayıp veri yöntemleri ile tamamlanmış ve kovaryans matris hesaplanmış ve bu kovaryans matrislerden

standart formüller kullanılarak alfa katsayısı hesaplanmıştır. Kayıp veri yöntemleri yanlılık, hataların ortalama karekökü ve güven aralığı bağlamında incelenmiştir. TROK mekanizmasında üç kayıp veri yöntemi tarafsız bulunmuş ve OA yöntemi negatif taraflı güvenilirlik tahminleriyle sonuçlanmıştır. ROK mekanizmasına sahip simülasyon verilerinde, yalnızca BM algoritması kovaryans matrisinden hesaplanan güvenilirlik tahminleri yansız bulunmuştur. Hatanın ortalama karekökü açısından kayıp değere ortalamayı atamak evrene ilişkin değerden daha düşük dağılıma sebep olmuştur. Bu çalışmada BM yaklaşımının (a) güvenilirlik tahminlerinde daha az yanlılık sağladığını, (b) tahminler bakımından örneklem arası dalgalanmayı önemli ölçüde azalttığını ve (c) daha kesin güven aralıklarını verdiğini göstermiştir. Mevcut sonuçlar, kayıp değerlerin güvenilirlik tahminlerinin doğruluğunu ve etkisini ve güven aralıklarını olumsuz yönde etkileyebileceğini göstermektedir. Araştırmacılara, veri matrisindeki eksik değerlerin sayısını açıkça tartışmaları ve geleneksel kayıp veri yöntemleri yerine EÇO yaklaşımını tercih etmeleri tavsiye edilmiştir.

Shrive, Stuart, Quan ve Ghali (2006), yaptıkları çalışmada depresyon ölçeğindeki kayıp verileri altı farklı atama yöntemiyle tamamlayıp karşılaştırmışlardır. Çalışmanın amacı kendi kendine raporlama (self-reporting) tekniğindeki ölçeklerde sıkça karşılaşılan kayıp veri ile baş etmek için alan yazındaki yöntemleri karşılaştırmaktır. Ölçek, katılımcıların her madde için 1 ile 4 arası bir değeri seçtiği 20 maddeden oluşan bir depresyon ölçeğidir. 1580 katılımcının olduğu çalışmada veri setindeki kayıp değerleri tamamlamak için ÇA, RA, BOA, genel ortalama atama, katılımcının bir önceki yanıtını atama, 1 ile 4 arası rastgele değer atama yöntemleri kullanılmıştır. Her bir yöntem için Spearman korelasyon katsayısı ve Kappa istatistiği hesaplanmıştır. Değerlerin %10 u eksik olduğunda rastgele değer atama dışındaki tüm yöntemler mükemmel yakın olarak nitelendirilen .80'den büyük Kappa istatistiği üretmiştir. ÇA bu senaryoda en iyi istatistiği vermiştir. Kayıp değerler %30'a ulaştığında ya da dengesiz kayıp değerlerin olduğu durumlarda ise ÇA yüksek Kappa istatistik değerini korumuştur.

Van Ginkel (2007), altı farklı kayıp veri baş etme yöntemini Cronbach alfa katsayısı, Loevinger'in ölçeklenebilirlik katsayısı H ve Mokken'in madde kümeleme katsayısını baz alarak simüle edilmiş 100 farklı veri seti üzerinden karşılaştırmalı

olarak incelenmiştir. İnceleme varyans analizi ile yapılmıştır. Varyans analizinde tam veri setlerinden elde edilen Cronbach alfa katsayıları arasındaki farklar bağımlı değişken, manipüle edilen koşullar ise bağımsız değişken olmuştur. Analizin çıktılarına göre %5 kayıp veri oranı için hata terimi içeren iki yönlü atama ve düzeltilmiş madde ortalaması atama teknikleri en ideal kayıp veri yöntemi olarak tespit edilmiştir. Kayıp veri oranı %15'e çıktığında ise rastgele değer atama haricindeki tekniklerin tümünün benzer sonuçlar verdiği görülmüştür. ÇA tekniği tüm koşullarda negatif yönlü sonuçlara sebep olmuştur (Akt. Akbaş, 2014).

Çokluk ve Kayrı (2011), kayıp değerlerin olmadığı veri setleri ile kayıp değerlere yaklaşık değer atama yöntemlerinin kullanıldığı veri setlerinin geçerlik ve güvenilirliğini karşılaştırmalı olarak incelemiştir. Çalışmanın amacı, kayıp değer olmadığı ve farklı oranlarda kayıp değere sahip veri setlerinde faktör yapılarının, düzeltilmiş madde-toplam korelasyonlarının ve iç tutarlılık katsayılarının hesaplanması ve karşılaştırılmasıdır. 200 kişiden elde edilen tam veri seti %15-%20 ve %0-%50 aralığında eksiltiyle kayıp değerlerin rastgele dağıldığı veri setleri elde edilmiştir. Araştırmada yaklaşık değer atama yöntemlerinden; Seriler Ortalaması Atama, Yakın Noktaların Medyanını Atama, Doğrusal Değer Kestirimi ve Noktanın Doğrusal Eğimi dikkate alınmıştır. Yapılan analizler sonucunda orijinal veri setindeki tek faktörlü yapının yaklaşık değer atama yöntemleri ile oluşturulmuş veri setlerinde de korunduğu görülmüştür. Önemli bir çıktı olarak bu yöntemler, açıklanan varyans oranlarında, faktör yüklerinin öz değerlerinde ve iç tutarlılığa yönelik hesaplanan Cronbach-alfa katsayısında düşüşe sebep olmuşlardır.

Misztal (2012), R paket programında uygulanan seçili kayıp veri atama yöntemlerini gözden geçirmek amacıyla bir çalışma yapmıştır. Öncelikle kayıp değerler üzerinde LBS yöntemi uygulanmış sonrasında ise kayıp değerlerin R paket programında uygulanan farklı atama yöntemleri kullanılarak nasıl ele alınacağını gösteren bir örnek sunulmuştur. Tüketici kredisi verilen 467 kişilik veri setinde borçluları iki risk sınıfına ayırmak için lojistik regresyon modeli temelinde analiz yapılmıştır. 467 kişiden elde edilen eksiksiz veri setinden, değerlerin %5.72'si rastgele kaldırılmış ve eksik veri seti elde edilmiştir. Elde edilen kayıp verili data atama yöntemleri ile tamamlanmış ve analiz yapılmıştır. Çalışmada örnek sunulduğu için genel sonuçlara ulaşmak mümkün olmamış fakat en kötü sonuç LBS

yönteminden elde edilmiştir. İlgili çalışmalardan da yola çıkarak en iyi performansın ÇA yöntemleri ile sergilendiği belirtilmiştir.

Demir (2013), kayıp değerler içeren ve iki kategorili puanlanan veri setlerinde kayıp veri yöntemlerini testlerin psikometrik özellikleri bağlamında incelemiştir. Bu özellikteki testlerde hangi yöntemin daha avantajlı olduğunu bulmayı amaçlamıştır. LBS, BA, EÇO ve ÇA yaklaşımına dayalı 12 yöntem çalışmaya konu olmuştur. Parametrelerin kestirimine yönelik yapılan analizler sonucunda, bu tür testlerde SA ve LBS yönteminin uygun olmadığı ve BA yöntemlerinin yanlış sonuçlar ürettiği görülmüştür. EÇO ve ÇA yöntemlerinin iki kategorili puanlanan ihmal edilemeyen kayıp veri içeren testler için uygun olduğu gözlenmiştir.

Başman (2014), kayıp verilere değer atamada kullanılan yapay sinir ağları yönteminin likert tipi ölçeklerde kullanımını incelemiştir. Çalışmada 2009 Pisa Türkiye örnekleminde okumaya ilişkin tutuma yönelik 11 maddelik ankete ait yanıtlar kullanılmıştır. 500 kişiye ait yanıtların yer aldığı eksiksiz veri seti %5, %10, %20 oranlarında eksiltilmiştir. Elde edilen kayıp değerli veri setleri yapay sinir ağları yöntemi ile tamamlanmıştır. Tüm veri setlerine açımlayıcı faktör analizi ve güvenilirlik analizi yapılmıştır. Tam veri setinden elde edilen değerler ile kayıp veri yöntemiyle tamamlanmış veri setinden elde edilen değerler karşılaştırılmıştır. Elde edilen bulgular, tam veri setinin tek faktörlü yapısının korunduğunu göstermiştir. Cronbach alfa katsayısı, omega katsayısı, açıklanan varyans oranı ve düzeltilmiş madde toplam korelasyonları değerlerinin tam veri seti değerleri ile aynı kaldığı görülmüştür.

Weaver ve Maxwell (2014), SPSS kullanıcıları için, veri setinde kayıp değer olması durumunda açımlayıcı faktör analizi ve güvenilirlik analizi üzerine etkisini incelemişlerdir. SPSS’de sıkça karşılaşılan LBS, ÇBS ve OA yöntemlerinin ilgili analizler üzerindeki yanlılığına vurgu yapılmış ve BM yöntemi temelinde bir uygulama ele alınmıştır.

Zhu (2014), boylamsal çalışmalarda sıklıkla karşılaşılan kayıp veri ile baş etmek amacıyla kullanılacak olan dört farklı kayıp veri yöntemini parametre tahminleri bağlamında karşılaştırmıştır. Kullanılan yöntemler, OA, LBS, gözlenen son değeri atama ve ÇA dır. Kayıp veri baş etme yöntemleri, kayıp veri oranı, kayıp veri mekanizması ve zamanın etkisi gibi değişkenler manipüle edilmiştir.

Yöntemlerin etkililiğın daha iyi yorumlayabilmek için yanlılık, RMSE ve %95 kapsama olasılıđı deęerleri hesaplanmıřtır. Elde edilen bulgular incelendiđinde gözlenen son deęeri atama yöntemi farklı kořullar altında en büyük yanlılıđa sebep olmuřtur. Veriler TROK mekanizması altındayken ÇA ve LBS benzer yanlılık ve %95 kapsama olasılıđı deęerlerini vermiřlerdir. Fakat veriler ROK iken ÇA en küçük yanlılık ve RMSE deęerine sahiptir. Bu nedenle TROK mekanizmasındaki veri setlerinde ÇA ve LBS yöntemleri tavsiye edilirken, ROK veri setlerinde ÇA yönteminin daha güvenilir bir yöntem olduđu sonucuna varılmıřtır.

Kürřad ve Nartgün (2015), kayıp veri sorununun çözümine iliřkin kullanılan yöntemlerin etkililiđini ölçeklerin psikometrik özellikleri bađlamında incelemiřlerdir. Bu amaçla 'Matematik Çalıřma Etiđi' ölçeđinden elde edilen Pisa 2012 Türkiye örnekleminde 200 kiřilik rastgele tam veri seti alınmıř ve TROK mekanizması altında farklı oranlarda kayıp veri setleri elde edilmiřtir. Bu kayıp veriler seri ortalaması, yakın noktaların ortalaması, yakın noktaların medyanı, dođrusal deęer kestirimi, noktanın dođrusal eđimi, LBS, BM, RA ve ÇA yöntemleri ile tekrar tam veri setine dönüřtürölmüřtür. Orijinal ve kayıp veri ile tamamlanmıř veri setlerine geçerlik ve güvenilirliđe iliřkin analizler uygulanmıřtır. Orijinal veri setinden elde edilen deęerler baz alınarak karřılařtırma yapılmıřtır. Yapılan karřılařtırmalar sonucunda LBS yöntemi referans deęerlere en az benzerlik gösteren yöntem olmuřtur. Yaklařık deęer atama yöntemleri düşük kayıp veri oranlarında tam veri setlerinden elde edilen sonuçlara yakın veya aynı sonuçları verirken, kayıp deęer arttıka elde edilen deęerler tam veri setinden elde edilen deęerlerden daha düşük bulunmuřtur. Tüm kayıp veri oranlarında en etkili sonuç veren yöntemler ise BM, RA ve ÇA olmuřtur.

Akbař ve Tavřancıl (2015), kayıp veri yöntemlerini, ölçeklerin psikometrik özellikleri bađlamında farklı kořullar altında incelemiřlerdir. Örnekleme büyüklüđu, kayıp veri oranının yanı sıra kayıp veri örüntüsü ve madde sayısı da farklılařtırılarak çalıřma geniřletilmiřtir. Bunun için 100 farklı yapay veri seti simöle edilmiřtir. Veri setleri farklı kayıp veri örüntülerinde eksiltilmiř ve LBS, Öklid uzaklıđı üzerinden benzer tepki örüntüsüne dayalı atama, stokastik regresyonla deęer atama, BM ve ÇA yöntemleri ile tamamlanmıřtır. Yapılan analizlerin çıktılarına göre LBS yönteminin tamamıyla rastgele kayıp mekanizmasına ve düşük kayıp veri oranına sahip veri

setleri hariç kullanılmaması gerektiği belirtilmiştir. ÇA yöntemi diğer yöntemlere göre açık bir üstünlük göstermiş fakat tüm durumlarda kullanılabilir ve en etkili sonuç veren bir yöntem bulunamamıştır.

Karaman, Doğan ve Soysal (2017), kayıp veri atama yöntemlerinin G kuramı G ve Phi katsayısına olan etkisini incelemek için bir çalışma yapmışlardır. Çalışmada birey (b) ve madde (m) olmak üzere $b \times m$ çaprazlanmış desen kullanılmıştır. Zayıf ve güçlü tek boyutluluk desenindeki normal dağılım altındaki veri setlerinde örneklem büyüklüğü (100, 200, 400, 1000) ve kayıp veri oranı (%5, %10, %20, %30) manipüle edilerek sıfır atama (SA), RA, OA, ÇA yöntemlerinin karşılaştırmalı olarak incelenmesi amaçlanmıştır. Tam veri setinden elde edilen değerler farklı koşullardaki veri setlerinden elde edilen değerler ile karşılaştırılmıştır. Karşılaştırmanın daha iyi yapılabilmesi için hataların kareleri ortalamasının karekökü (RMSE) ve yanlılık değerleri hesaplanmıştır. Çalışmanın sonunda SA yönteminin yanlı olacağı, RA ve ÇA yöntemlerinin kullanılmasının uygun olacağı, zayıf tek boyutluluk deseninin güçlü tek boyutluluk desenine göre daha yanlı olacağı ve örneklem büyüklüğü artışının hata ve yanlılık değerlerinde artışa neden olacağı öngörülmüştür.

Arıkan ve Soysal (2018), kayıp veri yöntemlerini farklı kayıp veri koşulları altında güvenilirlik katsayıları kestirimleri bağlamında karşılaştırmalı olarak incelemiştir. Örneklem büyüklüğü (500 ve 5000), kategori sayısı (3 ve 5) ve kayıp veri oranı (%5, %10, %20 ve %30) araştırmada manipüle edilmiştir. Kullanılan kayıp veri yöntemleri SA, RA, OA ve ÇA'dır. Cronbach alfanın yanında Standartlaştırılmış alfa, Armor'un θ , Guttman λ_4 , Guttman λ_5 , Guttman λ_6 ve McDonald'un ω güvenilirlik katsayılarının kestirimleri yapılmış ve tam veri setinden elde edilen değerler referans değer olarak kullanılmıştır. Analizin çıktılarına göre güvenilirlik katsayıları için en iyi olarak kabul edilebilecek kayıp veri yönteminin olmadığı görülmüştür. Bunun yanında ÇA ve RA'nın tercih edilmesi gerektiği, SA yönteminden uzak durulması gerektiği belirtilmiştir.

Bayhan (2018), iç tutarlığın kayıp veri oranından ne derece etkilendiğini incelemiştir. Cronbach alfa katsayısı; farklı oranlardaki kayıp veri oranının (%5, %10, %20) yanında örneklem büyüklüğü (100, 250, 500, 1000), test uzunluğu (5, 10, 25, 50), dağılım biçimi (normal dağılım, sağa çarpık dağılım, sola çarpık dağılım) ve puanlama türü (kategorik ve likert) değişkenleri açısından da incelenmiştir. Analiz

sonucunda her iki puanlama türünde de örneklem büyüklüğünün artışı güvenilirlik katsayısındaki değişimin miktarını azaltmıştır. Aynı şekilde madde sayısının artışı Cronbach alfa katsayısındaki değişim miktarını azaltmıştır. Elde edilen bulgularda dağılım biçiminin güvenilirlik katsayısına etkisinin olmadığı görülmüştür. Cronbach alfa değerindeki değişim; örneklem büyüklüğü, test uzunluğu ve dağılım biçimi farketmeksizin kayıp veri oranına bağlı olarak artmaktadır.

Soysal, Karaman ve Doğan (2018), farklılaşan kayıp veri oranı (%5, %10, %20, %30) ve örneklem büyüklüğü (100, 200, 400, 1000) koşulları altında kayıp verinin G ve Phi katsayısına etkisini araştırmışlardır. Zayıf ve güçlü tek boyutlu normal dağılım altındaki veri setleri belirtilen oranlarda eksiltilmiş ve SA yöntemiyle tamamlanmıştır. Tam veri seti, kayıp verili set ve SA yapılmış veri setinden elde edilen katsayı değerleri karşılaştırılmıştır. Daha iyi yorum yapabilmek için yanlılık ve hataların kareleri ortalamasının karekökü (RMSE) değerleri hesaplanmıştır. Analiz çıktılarına göre, zayıf tek boyutlu desenler kayıp veri oranının %20 ve daha fazla olduğu durumlarda etkilenmiş fakat güçlü tek boyutlu desenindeki veri setleri %30 oranında bile minimal düzeyde etkilenmişlerdir. Zayıf tek boyutlu veri setlerindeki hata değerleri daha fazla bulunmuştur. SA yöntemiyle elde edilen güvenilirlik kestirimleri daha yanlı bulunduğu için bu yöntemin kullanılmaması, güvenilirlik çalışmalarında kayıp değer bulunan veri setleriyle çalışılmasına imkan sağlayan G kuramına yer verilmesi önerilmiştir.

De Raadt, Warrens, Bosker ve Kiers (2019), iki puanlayıcı arasındaki karşılaştırmalı uyuşmanın güvenilirliğini ölçen bir istatistik yöntem olan Cohen'in Kappa katsayısını veri setinde kayıp değer olması durumunda incelemişlerdir. TROK ve İEK veri mekanizması altında Kappa katsayısının ne kadar iyi tahmin edildiğini araştırmışlardır. Kayıp olan verileri tamamlamak için biri LBS olan üç farklı yöntem kullanmışlardır. Yapılan çalışmanın sonucunda eksik verileri silinen TROK mekanizmasına sahip veri setinden elde edilen Kappa katsayısında yanlılık ve hata bulunmamıştır. Yine eksik verileri silinen fakat İEK veri mekanizmasındaki veri setinden elde edilen katsayılar düşük miktarda hata ve yanlılık bulunmuştur. Yüksek performans gösterdiği ve analizi kolay olduğu için her iki kayıp veri mekanizmasında da LBS yönteminin kullanılması önerilmiştir.

3. YÖNTEM

Bu çalışmada araştırmanın modeli, çalışma verileri, verilerin analiz yöntemleri ele alınmıştır.

3.1 Araştırmanın Modeli

Bu araştırmada, farklı kayıp veri ile baş etme yöntemleri ile tamamlanmış veri setleri üzerinde G kuramı ile güvenilirlik kestirimi çalışması yapılmıştır. TROK mekanizmasına sahip %5, %10 ve %20 oranlarında eksik değere sahip veri setleri kayıp veri yöntemleri kullanılarak tamamlanmıştır. Kayıp veri yöntemlerinden LBS, OA ve RA yöntemleri kullanılmıştır. Çalışmada, $b \times m \times p$ çaprazlanmış tesadüfi desen kullanılmıştır. Birey (b), madde (m) ve puanlayıcı (p) değişkenlik kaynaklarıdır. Yapılan G çalışması ile tam veri ve kayıp veri yöntemi ile tamamlanmış veri seti üzerinde test için değişkenlik kaynaklarına ait varyans bileşenleri hesaplanmıştır. K çalışması ile elde edilen varyans bileşenleri kullanılarak hesaplanan G ve Phi katsayılarının nasıl farklılaştığı incelenmiştir. G kuramı çok değişkenli G çalışması ile elde edilen varyansların ve K çalışması ile elde edilen G ve Phi katsayılarının hesaplanıp hangi durumda analizin nasıl sonuçlar ürettiği karşılaştırıldığı için çalışma kuramsal bir araştırmadır.

3.2 Çalışma Verileri

Bu çalışmada 689 kişi, bir test ve dört puanlayıcının olduğu gerçek bir ölçme durumundan elde edilen bir ana veri matrisi kullanılmıştır. Gerçek ölçme durumu yedi boyut ve bu boyutlara ait toplamda 28 görevden oluşmaktadır. Puanlar 1-0 olup kategoriktir. Her bir boyuta ait puanlar toplanmış ve puanlar sürekli hale getirilmiştir. Tam veri matrisi yedi maddeden oluşan bir teste dönüştürülmüştür. Tüm bireyler tüm maddeleri yanıtlamışlar ve tüm puanlayıcılar tüm bireyleri tüm maddelerde puanlamışlardır. Aynı zamanda madde ve puanlayıcı yüzeyleri (facet) random olarak alınmıştır. Bu nedenle veri matrisi çaprazlanmış random bir desen oluşturur.

Literatürde G kuramı ve güvenilirlik ile ilgili çalışmalar incelendiğinde farklı

örneklem büyüklükleri tespit edildi. Enders (2004), güvenilirlik çalışmasını örneklem büyüklüğünü 100, 300 ve 500 olarak gerçekleştirmiştir. Çokluk ve Kayrı (2011), 200 kişilik örneklem üzerinde, Doğan ve Soysal (2017) ve Soysal, Karaman ve Doğan (2018) 100, 200, 400 ve 1000 örneklem büyüklüğü ile analizlerini gerçekleştirmişlerdir. Nunnally ve Bernstein (1994), güvenilirlik üzerine yapılan çalışmalarda hata miktarını azaltmak için örneklem büyüklüğünün en az 300 olması gerektiğini belirtmişlerdir. Atılgan (2013), örneklem büyüklüğü 30 olduğunda G ve Phi katsayılarının yeterince tarafsız tahmin edilemeyeceğini, 50 ve 300 arası örneklem büyüklüğünde tarafsız tahmin yapılabildiğini belirtmiştir. 400 ve üzeri örneklem büyüklüğünde elde edilen RMSE değerlerinin sabit olduğunu, kesin ve kararlı G ve Phi katsayısı tahmini için 400 örneklem büyüklüğünün yeterli olduğunu söylemiştir. Çalışmalar göz önüne alındığında gerçek ölçme durumundaki 689 kişilik ana veri matrisi yeterli görülmüştür.

Alan yazın incelendiğinde kayıp veri oranı ve mekanizmasının farklı oranlarda ve şekillerde alındığı görülmüştür. Köse ve Öztemur (2014), Kürşat ve Nartgün (2015) ve Başman (2014), %5, %10 ve %20 oranlarında kayıp değere sahip tamamen seçkisiz mekanizmasına sahip verilerle çalışmalarını gerçekleştirmişlerdir. Shrive, Stuart, Quan ve Ghali (2006), %10 ve %30 kayıp veri oranının etkilerini incelemişlerdir. Çokluk ve Kayrı (2011), seçkisiz kayıp veri mekanizmasında %15-%20 arası ve %0-%50 arası kayıp değere sahip iki adet veri matrisi kullanmışlardır. Enders (2003), TROK, ROK ve İEK koşulları altında %15 ve %30 oranlarında kayıp değere sahip verileri kullanmıştır. Soysal, Karaman ve Doğan (2018), %5, %10, %20 ve %30 oranlarında tamamen seçkisiz kayıp verileri kullanmışlardır. Tabachnick ve Fidell (2013), %5 ve daha az kayıp değere sahip tamamen seçkisiz kayıp verilerin ihmal edilebileceğini belirtmiştir. Çalışmada kullanılan kayıp veri yöntemleri dikkate alınarak alan yazın incelendiğinde kayıp veri oranı % 20 nin üzerinde olduğunda tamamlanan veri setinin ana veri setinden büyük oranda farklılaştığı görülmektedir. Bu da baz alınan diğer koşullardan bağımsız olarak hata miktarını arttırmaktadır. İncelenen çalışmalar göz önüne alınarak R’da yazılan kodlarla ana veri matrisi %5, %10 ve %20 oranlarında tamamen seçkisiz olarak eksiltilerek normal dağılıma sahip matrisler elde edilmiştir. Kayıp değer içeren veriler için ana veri matrisinde olduğu gibi birey, madde ve puanlayıcı (*bmxmp*) çaprazlanarak yeni

matrisler oluşturulmuştur. Bağımlı değişken olan toplam puan istenilen kayıp veri oranı baz alınarak tamamen seçkisiz olarak silinmiştir.

Araştırmada kullanılan veri matrisleri simülasyon ile elde edilmiştir. Simülasyon sosyal bilimlerde sosyal olguları süreçler bağlamında anlamının bir yolu olarak kullanılmaya başlanan bilgisayar ortamında gerçekleştirilen model deneylerdir. Bir fikir olarak bilgisayar simülasyonu, bilgisayarların başlangıcına kadar uzanan bir geçmişe sahiptir ancak sosyal bilimlerde yaygın kullanımı 1990 ların başında gerçekleşmiştir. Simülasyon sayesinde sosyal kavramlar ve süreçler hakkında teori kurmayı kolaylaştıran bilgisayar modelleri kurulmuş, farklı değişkenlerin modeli nasıl etkileyeceği test edilmiş, tek bir sonuç yerine olası senaryolar elde etmek ve model geliştiriciden başka birinin pratik bir amaç için kullanabileceği model sunmak mümkün hale gelmiştir (Gilbert, 1999).

Tekrar (replikasyon), kümülatif bilimin işaretlerinden biridir. Belirli bir simülasyon modelinin iddia edilen sonuçlarının sıfırdan başlanarak bir başkası tarafından üretilebilmesi açısından güvenilir olup olmadıklarının doğrulanması gerekir. Bu onay olmadan, yayınlanan bazı sonuçların programlama hataları, sonuçları analiz etme veya raporlamadaki hatalardan dolayı yanlış olması mümkündür. Tekrarlama, modellerden çıkarımların sağlamlığını test etmek açısından önemlidir (Axelrod, 1997). Çalışmada ana veri matrisinin özelliklerinin dikkate alındığı 100 adet veri matrisi simule edilmiştir.

Kayıp verili matrisler daha sonra LBS, OA ve RA yöntemleri ile tekrardan tam veri setine dönüştürülmüştür. Atama yöntemleri için R programından *mice* (Buuren & Groothuis-Oudshoorn, 2011) ve *norm* (Novo & Schafer, 2013) paketleri kullanılmıştır. Araştırmada üç farklı kayıp veri oranı (%5, %10 ve %20), üç farklı kayıp veri yöntemi (LBS, OA ve RA) olmak üzere 3x3 adet desen elde edilmiştir. Her bir desen için 100 tekrar yapılmış olup 9x100 olmak üzere toplamda 900 adet veri matrisi kullanılmıştır.

3.3 Verilerin Analizi

Verilerin analizinde başlangıçta tam veri setinde G ve K çalışmaları yapılmıştır. G çalışması ile elde edilen varyans bileşenleri kullanılarak G ve Phi

katsayıları hesaplanmıştır. Çalışmada ayrıca evren puanı varyansı, mutlak hata varyansı ve bağıl hata varyansı da dikkate alınmış ve yorumlanmıştır. Daha sonra kayıp veri yöntemleri ile tamamlanmış toplamda 100 tekrar üzerinde de aynı analizler yapılmıştır. Her bir kayıp veri oranı ve kayıp veri yöntemi için 100 tekrar üzerinden elde edilen G katsayıları, Phi katsayıları, evren puanı varyansları, mutlak hata varyansları ve bağıl hata varyanslarının ortalaması alınmıştır. Verilerin analizinde R yazılımından *gtheory* (Moore, 2016) paketi kullanılmıştır.

Evren puanı klasik test kuramındaki gerçek puan ile benzerlik gösteren, bireyin ortalama puanının beklenen değeridir. Evren puanı varyansı evren puanının tüm bireyler üzerindeki varyansıdır. Bağıl hata varyansı birey ve diğer yüzeylerin ortak etki varyanslarından oluşur. Birey ve diğer yüzeylerin etkileşimi ile oluşan varyans bileşenleri σ^2_{bm} , σ^2_{bp} ve $\sigma^2_{bmp,e}$ olur. Bu sebeple bağıl hata varyansı aşağıdaki gibi elde edilir:

$$\sigma^2(\delta) = \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{bmp,e}}{n_m n_p} \quad (10)$$

Mutlak hata varyansı birey hariç tüm yüzeylerin ortak ve ana etki varyanslarını kapsar. İki yüzeyli çaprazlanmış desende elde edilen birey dışındaki varyans bileşenlerinin (σ^2_m , σ^2_p , σ^2_{bm} , σ^2_{bp} , σ^2_{mp} ve σ^2_{bgp}) toplamından oluşur. Mutlak hata varyansı aşağıdaki gibi elde edilir:

$$\sigma^2(\Delta) = \frac{\sigma^2_m}{n_m} + \frac{\sigma^2_p}{n_p} + \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{mp}}{n_m \cdot n_p} + \frac{\sigma^2_{bmp,e}}{n_m \cdot n_p} \quad (11)$$

Genellenebilirlik katsayısı ($E\rho^2$) evren puan varyansının kendisi ve bağıl hata varyansı toplamına oranıdır ve aşağıdaki gibi gösterilir:

$$E\rho^2 = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2(\delta)} \quad (12)$$

Güvenilirlik indeksi (ϕ) evren puanı varyansının kendisi ve mutlak hata varyansı toplamına oranıdır ve aşağıdaki gibi gösterilir:

$$\phi = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2(\Delta)} \quad (13)$$

Analiz sonuçları tam veri setinden ve kayıp veri yöntemleri ile tamamlanmış verilerden elde edilen değerler karşılaştırılarak değerlendirilmiştir. Farklı oranlarda eksiltelen veri setleri ile tam veri setlerinden elde edilen istatistikler arasındaki sapmalar RMSE (Root Mean Square Error) ve yanlılık ile incelenmiştir. RMSE, gerçek ve kestirilen parametre değerleri arasındaki farkın karesinin ortalamasının kareköküdür. Araştırma kapsamında hesaplanan RMSE değerine ait matematiksel Eşitlik 14'teki gibi gösterilir:

$$\text{RMSE} = \sqrt{\frac{(\bar{\tau}_j - \tau_j)^2}{R}} \quad (14)$$

$\bar{\tau}_j$: j parametresinin gerçek değeri

τ_j : j parametresinin kestirilen değeri

R: Replikasyon sayısı

Dokuz desenin her biri için G katsayısı, Phi katsayısı, evren puanı varyansı, mutlak hata varyansı ve bağıl hata varyansı istatistiklerinin RMSE değeri hesaplanmıştır.

Yanlılık kestirilen parametre değerlerinin ortalamasının gerçek parametre değerinden farkıdır. Yanlılık değeri Eşitlik 15'te görüldüğü gibi hesaplanır:

$$\text{Yanlılık} = \frac{(\sum_{n=1}^R \bar{\tau}_{nj})}{R} - \tau_j \quad (15)$$

$\bar{\tau}_{nj}$: j parametresinin n. tekrara ait kestirilen değeri

τ_j : j parametresinin gerçek değeri

R= Replikasyon sayısı

Dokuz desenin her biri için G katsayısı, Phi katsayısı, evren puanı varyansı, mutlak hata varyansı ve bağıl hata varyansı istatistiklerinin yanlılık değeri hesaplanmıştır.

4. BULGULAR

Bu bölümde araştırmanın amaçları doğrultusunda G kuramı çok değişkenli analiz sonuçlarına yer verilmiştir. Elde edilen bulgular ve yapılan yorumlar alt problemlerin sırasına göre yapılmıştır.

4.1 Birinci Alt Probleme Ait Bulgu ve Yorumlar

Aynı veri setinden %5 oranlarla oluşturulan eksik verili veri setlerinde;(a) “LBS” (b) “OA” ve (c) “RA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

Tam veri seti ve %5 oranla oluşturulan eksik verili veri setlerinden elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısına ait ortalamalar Tablo 1'de gösterilmiştir.

Tablo 1. Tam Veri Seti ve %5 Oranla Oluşturulan Eksik Verili Veri Setlerinden Elde Edilen Evren Puanı Varyansı, Bağıl Hata Varyansı, Mutlak Hata Varyansı, G ve Phi Katsayısı Ortalamaları

		Evren Puanı Varyansı	Bağıl Hata Varyansı	MutlakHata Varyansı	G	Phi
	Tam Veri Seti	.384	.140	.291	.732	.569
%5 Eksik Veri	LBS	.384	.141	.291	.731	.569
	OA	.347	.131	.267	.726	.565
	RA	.399	.128	.279	.757	.589

LBS: Liste Bazında Silme; OA: Ortalama Atama; RA: Regresyon Atama

Tablo 1'e bakıldığında %5 oranla oluşturulan eksik verili veri setlerinde LBS yönteminin kullanımıyla elde edilen evren puanı varyansı, mutlak hata varyansı ve Phi katsayısı ortalamaları tam veri setinden elde edilen istatistiklerle aynıdır. Diğer yandan %5 oranla oluşturulan eksik verili veri setlerinde LBS yönteminin kullanımıyla elde edilen bağıl hata varyansı ve G katsayısı için elde edilenden farkı sırasıyla .001 ve -.001'dir. %5 oranla oluşturulan eksik verili veri setlerinde OA

yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı ortalamalarının tam veri setinden elde edilen istatistiklerden farkı sırasıyla -.037, -.009, -.024, -.006, -.004 olarak hesaplanmıştır. %5 oranla oluşturulan eksik verili veri setlerinde RA yönteminin kullanımıyla elde edilen bağıl hata varyansı, mutlak hata varyansı ortalamalarının tam veri setinden elde edilen istatistiklerden farkı her ikisi için de -.012, evren puanı varyansı, G ve Phi katsayıları için farkı ise sırasıyla .015, .025 ve .02'dir.

%5 oranda eksik veri içeren LBS, OA ve RA yöntemleriyle tamamlanan veri setlerinde evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayıları için hesaplanan yanlışlık ve RMSE değerleri Tablo 2'de gösterilmiştir.

Tablo 2. %5 Oranla Oluşturulan Eksik Verili Veri Setlerinde Hesaplanan Hata İstatistikleri

Yöntem	Hata	Evren Puan Varyansı	Bağıl Hata Varyansı	Mutlak Hata Varyansı	G	Phi
LBS	RMSE	.000205	.000003	.000004	.000138	.000144
	Yanlışlık	-.001420	.004989	.004032	-.001720	-.002350
OA	RMSE	.141372	.009018	.059161	.004830	.002159
	Yanlışlık	-.091790	-.073570	-.086100	-.005340	-.002690
RA	RMSE	.022473	.014955	.014976	.059975	.038848
	Yanlışlık	.041961	-.090990	-.043610	.035358	.036684

LBS: Liste Bazında Silme; OA: Ortalama Atama; RA: Regresyon Atama

%5 eksik verili veri setlerinde LBS yöntemi kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayılarına ait RMSE değerleri .000003 ile .000205 arasında değişmektedir. Bu durum en yüksek değere sahip olan evren puanı varyansının en fazla hata ile kestirildiğini, en düşük değere sahip bağıl hata varyansının daha kararlı davrandığı şeklinde yorumlanabilir. Aynı eksik veri oranına sahip veri setlerinde OA yöntemi kullanımıyla elde edilen istatistiklere ait RMSE değerleri .002159 ile .141372 arasında değişmektedir. LBS yöntemine benzer şekilde OA yönteminde de evren puanı varyansı en fazla hata ile elde edilmiştir. RA yöntemi kullanılan veri setlerinde RMSE değerleri .014955 ile .059975 arasında saptanmıştır. Bu yöntemde en fazla hata ile kestirilen istatistiğin G katsayısı olduğu, en kararlı davranan istatistiğin bağıl hata varyansı olduğu

söylenbilir. Tüm kayıp veri yöntemlerine bakıldığında en yüksek RMSE değerlerinin genel olarak RA yönteminde olduğu görülmektedir. LBS yöntemi kullanılan veri setlerinde elde edilen istatistiklere ait yanlılık değerlerinin -.00235 ve .004989 arasında değiştiği görülmektedir. En yanlı istatistiğin bağıl hata varyansı olduğu görülmektedir. Mutlak değerce en küçük değere sahip evren puanı varyansı sıfıra yakın bir yanlılıkla kestirilmiştir. OA yöntemi kullanılan veri setlerinde istatistiklere ait yanlılık değerleri tüm istatistiklerde negatif değerlidir. En yanlı istatistik mutlak değerce en büyük olan evren puanı varyansıdır. Phi katsayısı sıfıra en yakın yanlılığa sahiptir. RA yöntemi kullanılan veri setlerinde kestirilen istatistikler -.04361 ve .041961 arasında değişen yanlılıkla kestirilmiştir. Tablo tümüyle incelendiğinde LBS yöntemine ait yanlılık değerlerinin sıfıra en yakın olduğu ve yansız kestirimler yapıldığı söylenebilir.

4.2 İkinci Alt Probleme Ait Bulgu ve Yorumlar

Aynı veri setinden %10 oranlarla oluşturulan eksik verili veri setlerinde; (a) “LBS” (b) “OA” (c) “RA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir?

Tam veri seti ve %10 oranla oluşturulan eksik verili veri setlerinden elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısına ait ortalamalar Tablo 3'te gösterilmiştir.

Tablo 3. Tam Veri Seti ve %10 Oranla Oluşturulan Eksik Verili Veri Setlerinden Elde Edilen Evren Puanı Varyansı, Bağıl Hata Varyansı, Mutlak Hata Varyansı, G ve Phi Katsayısı Ortalamaları

		Evren Puanı Varyansı	Bağıl Hata Varyansı	Mutlak Hata Varyansı	G	Phi
	Tam Veri Seti	.384	.140	.291	.732	.569
%10 Eksik Veri	LBS	.384	.142	.292	.730	.568
	OA	.312	.122	.244	.719	.561
	RA	.413	.116	.267	.780	.607

LBS: Liste Bazında Silme; OA: Ortalama Atama; RA: Regresyon Atama

Tablo 3'e bakıldığında %10 oranla oluşturulan eksik verili veri setlerinde LBS yönteminin kullanımıyla elde edilen evren puanı varyansı ortalaması tam veri setinden elde edilen istatistiklerle farkı sıfırdır. Bağlı hata varyansı, mutlak hata varyansı, G ve Phi katsayıları için ise fark sırasıyla .002, .001, -.002 ve -.001 olarak bulunmuştur. %10 oranla oluşturulan eksik verili veri setlerinde OA yönteminin kullanımıyla elde edilen evren puanı varyansı, bağlı hata varyansı, mutlak hata varyansı ile G ve Phi katsayılarının ortalamalarının tam veri setinden elde edilen istatistiklerden farkı sırasıyla -.072, -.018, -.047, -.013 ve -.008 olarak hesaplanmıştır. %10 oranla oluşturulan eksik verili veri setlerinde RA yönteminin kullanımıyla elde edilen bağlı hata varyansı, mutlak hata varyansı ortalamalarının tam veri setinden elde edilen istatistiklerden sapmaları her ikisi için de -.024 olarak hesaplanırken evren puanı varyansı, G ve Phi katsayıları için bu fark sırasıyla .029, .048 ve .038'dir. %10 oranda eksik veri içeren LBS, OA ve RA yöntemleriyle tamamlanan veri setlerinde evren puanı varyansı, bağlı hata varyansı, mutlak hata varyansı, G ve Phi katsayıları için hesaplanan yanlılık ve RMSE değerleri Tablo 4'te gösterilmiştir.

Tablo 4. %10 Oranla Oluşturulan Eksik Verili Veri Setlerinde Hesaplanan Hata İstatistikleri

Yöntem	Hata	Evren Puan Varyansı	Bağlı Hata Varyansı	Mutlak Hata Varyansı	G	Phi
LBS	RMSE	.000514	.000161	.000207	.000463	.000389
	Yanlılık	-.001120	.009485	.004904	-.00284	-.00260
OA	RMSE	.522555	.035277	.223908	.017630	.006942
	Yanlılık	-.188100	-.130530	-.16003	-.01867	-.01473
RA	RMSE	.085275	.057964	.057341	.229880	.147665
	Yanlılık	.075550	-.168180	-.080570	.064565	.066719

LBS: Liste Bazında Silme; OA: Ortalama Atama; RA: Regresyon Atama

%10 eksik verili veri setlerinde LBS yöntemi kullanımıyla elde edilen evren puanı varyansı, bağlı hata varyansı, mutlak hata varyansı, G ve Phi katsayılarına ait RMSE değerleri .000016 ile .000514 arasında değişmektedir. Elde edilen bu sonuç en yüksek değere sahip olan evren puanı varyansının en fazla hata ile kestirildiğini, en düşük değere sahip bağlı hata varyansının daha kararlı davrandığı şeklinde

yorumlanabilir. OA yöntemi kullanımıyla elde edilen istatistiklere ait RMSE değerleri .006942 ile .522555 arasında değişmektedir. Evren puanı varyansı en fazla hata ile kestirilmiştir. RA yöntemi kullanılan veri setlerinde RMSE değerleri .057341 ile .229880 arasında saptanmıştır. Bu yöntemde en kararlı davranan istatistiğin mutlak hata varyansı olduğu en fazla hata ile kestirilen istatistiğin ise %5 eksik verili veri setine benzer şekilde G katsayısının olduğu söylenebilir. Kayıp veri yöntemlerinin tamamına bakıldığında RA yönteminin en yüksek, LBS yönteminin en düşük RMSE değerlerine sahip olduğu görülmektedir. LBS yöntemi kullanılan veri setlerinde elde edilen istatistiklere ait yanlışlık değerlerinin -.001120 ve .009485 arasında değiştiği görülmektedir. En yanlış istatistiğin bağıl hata varyansı olduğu görülmektedir. Mutlak değerce en küçük değere sahip evren puanı varyansı sıfıra yakın bir yanlışlıkla kestirilmiştir. OA yöntemi kullanılan veri setlerinde istatistiklere ait yanlışlık değerleri tüm istatistiklerde negatif değerlidir. En yanlış istatistik mutlak değerce en büyük olan evren puanı varyansıdır. Phi katsayısı sıfıra en yakın yanlışlığa sahiptir. RA yöntemi kullanılan veri setlerinde kestirilen istatistikler -.080570 ve .075550 arasında değişen yanlışlıkla kestirilmiştir. G katsayısı mutlak değerce en küçük yanlışlık değerine sahiptir. En yanlış kestirim negatif yönlü olan mutlak hata varyansıdır. Tablo tümüyle incelendiğinde yansız kestirimlerin LBS yöntemine ait olduğu görülmektedir.

4.3 Üçüncü Alt Probleme Ait Bulgu ve Yorumlar

Aynı veri setinden %20 oranlarla oluşturulan eksik verili veri setlerinde; (a) “LBS” (b) “OA” ve (c) “RA” yönteminin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmaları nasıl değişmektedir.

Tam veri seti ve %20 oranla oluşturulan eksik verili veri setlerinden elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısına ait ortalamalar Tablo 5'te gösterilmiştir.

Tablo 5. Tam Veri Seti ve %20 Oranla Oluşturulan Eksik Verili Veri Setlerinden Elde Edilen Evren Puanı Varyansı, Bağlı Hata Varyansı, Mutlak Hata Varyansı, G ve Phi Katsayısı Ortalamaları

		Evren Puanı Varyansı	Bağlı Hata Varyansı	Mutlak Hata Varyansı	G	Phi
Tam Veri Seti		.384	.140	.291	.732	.569
%20 Eksik Veri	LBS	.385	.143	.294	.729	.567
	OA	.246	.104	.200	.703	.551
	RA	.440	.094	.245	.823	.641

LBS: Liste Bazında Silme; OA: Ortalama Atama; RA: Regresyon Atama

Tablo 5'e bakıldığında %20 oranla oluşturulan eksik verili veri setlerinde LBS yönteminin kullanımıyla elde edilen evren puanı varyansı, bağlı hata varyansı ve mutlak hata varyansı ortalamaları tam veri setinden elde edilen istatistiklerden farkı sırasıyla .001, .003 ve .003 olarak bulunmuştur. G katsayısı ve Phi katsayısı için sapmalar ise sırasıyla -.003 ve -.002'dir. %20 oranla oluşturulan eksik verili veri setlerinde OA yönteminin kullanımıyla elde edilen evren puanı varyansı, bağlı hata varyansı, mutlak hata varyansı, G ve Phi katsayısı ortalamalarının tam veri setinden elde edilen istatistiklerden farkı sırasıyla -.138, -.036, -.091, -.029 ve -.018 olarak hesaplanmıştır. %20 oranla oluşturulan eksik verili veri setlerinde RA yönteminin kullanımıyla elde edilen bağlı hata varyansı, mutlak hata varyansı ortalamalarının tam veri setinden elde edilen istatistiklerden farkı her ikisi için de -.046, evren puanı varyansı, G ve Phi katsayıları için ise fark sırasıyla .056,.091, .072 olarak bulunmuştur.

%20 oranda eksik veri içeren LBS, OA ve RA yöntemleriyle tamamlanan veri setlerinde evren puanı varyansı, bağlı hata varyansı, mutlak hata varyansı, G ve Phi katsayıları için hesaplanan yanlışlık ve RMSE değerleri Tablo 6'da gösterilmiştir.

Tablo 6. %20 Oranla Oluşturulan Eksik Verili Veri Setlerinde Hesaplanan Hata İstatistikleri

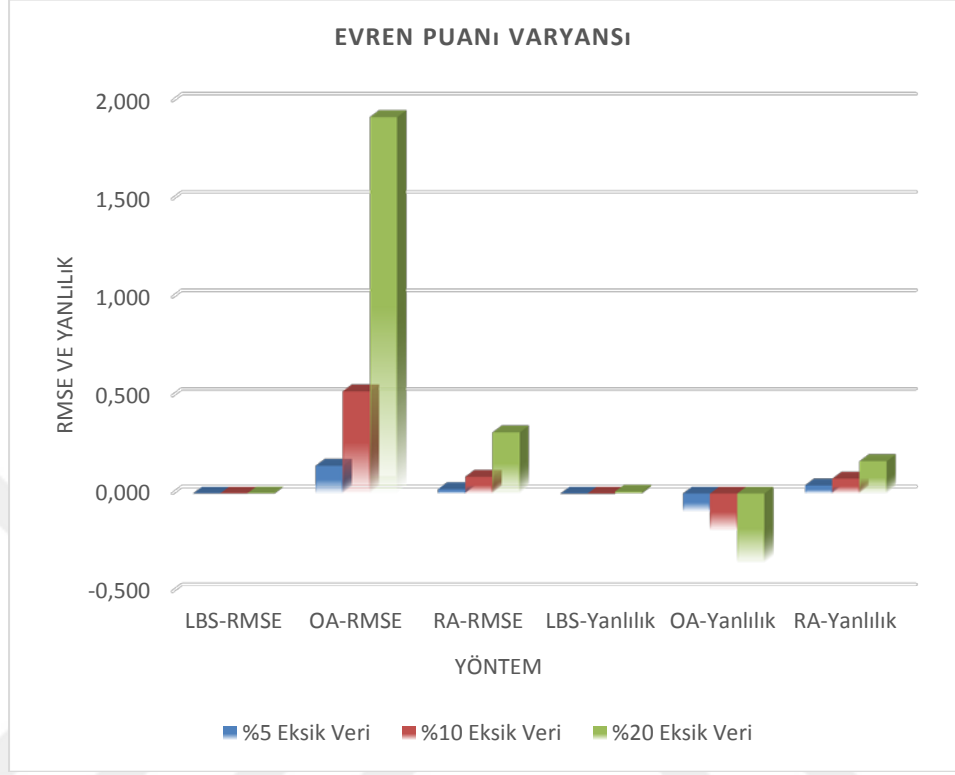
Yöntem	Hata	Evren Puan Varyansı	Bağlı Hata Varyansı	Mutlak Hata Varyansı	G	Phi
LBS	RMSE	.001152	.000742	.000875	.001726	.001103
	Yanlılık	.009668	.017321	.006055	-.002030	.001541
OA	RMSE	1.917185	.133266	.825107	.090034	.035139
	Yanlılık	-.354170	-.260160	-.314200	-.037550	-.026020
RA	RMSE	.311643	.210451	.207503	.822386	.529712
	Yanlılık	.164494	-.329380	-.158410	.128065	.135727

LBS: Liste Bazında Silme; OA: Ortalama Atama; RA: Regresyon Atama

%20 eksik verili veri setlerinde LBS yöntemi kullanımıyla elde edilen evren puanı varyansı, bağlı hata varyansı, mutlak hata varyansı, G ve Phi katsayılarına ait RMSE değerleri .000742 ile .001726 arasında değişmektedir. OA yöntemi kullanılan veri setlerinde kestirilen istatistiklere ait RMSE değerleri .035139 ile 1.917185 arasında değişmektedir. RA yönteminde ise RMSE .207503 ile .822386 arasında değer almıştır. Yöntemlerin tamamına bakıldığında; evren puanı varyansı ve mutlak hata varyansının OA yönteminde; bağlı hata varyansı ile G ve Phi katsayılarının RA yönteminde daha büyük hata ile kestirildiği görülmektedir. Buna karşın LBS yöntemi ile RMSE değerlerinin diğer yöntemlerden daha küçük olduğu bulunmuştur. LBS yöntemi kullanılan veri setlerinde elde edilen istatistiklere ait yanlılık değerleri -0,00203 ve 0,017321 arasında değişmektedir. En yanlı istatistiğin bağlı hata varyansı olduğu, yanlılığı en küçük olanın ise Phi katsayısı olduğu söylenebilir. OA yöntemi kullanılan veri setlerinde istatistiklere ait yanlılık değerleri diğer kayıp veri oranların da olduğu gibi tüm istatistiklerde negatif değerlidir. Yanlılık değerlerinin en büyüğü -.35417 ile evren puanı varyansı, en küçüğü -.02602 ile Phi katsayısıdır. RA yöntemi kullanılan veri setlerinde kestirilen istatistikler -.32938 ile.164494 arasında değişen yanlılıkla kestirilmiştir. G katsayısı için hesaplanan .12807 yanlılık değeri sıfıra en yakın yanlılık değeri olduğundan en küçük yanlılığa sahiptir. En yanlı kestirim ise sıfırdan en uzak olan -.32938 ile negatif yönlü olan bağlı hata varyansıdır.

İstatistiklerin RMSE ve yanlılık değerlerindeki değişimi kayıp veri yöntemleri ve kayıp veri oranları bazında daha iyi görebilmek için veriler grafik ile gösterilmiştir.

Evren puanı varyansı istatistiğine ait grafik Şekil 2 de yer almaktadır.

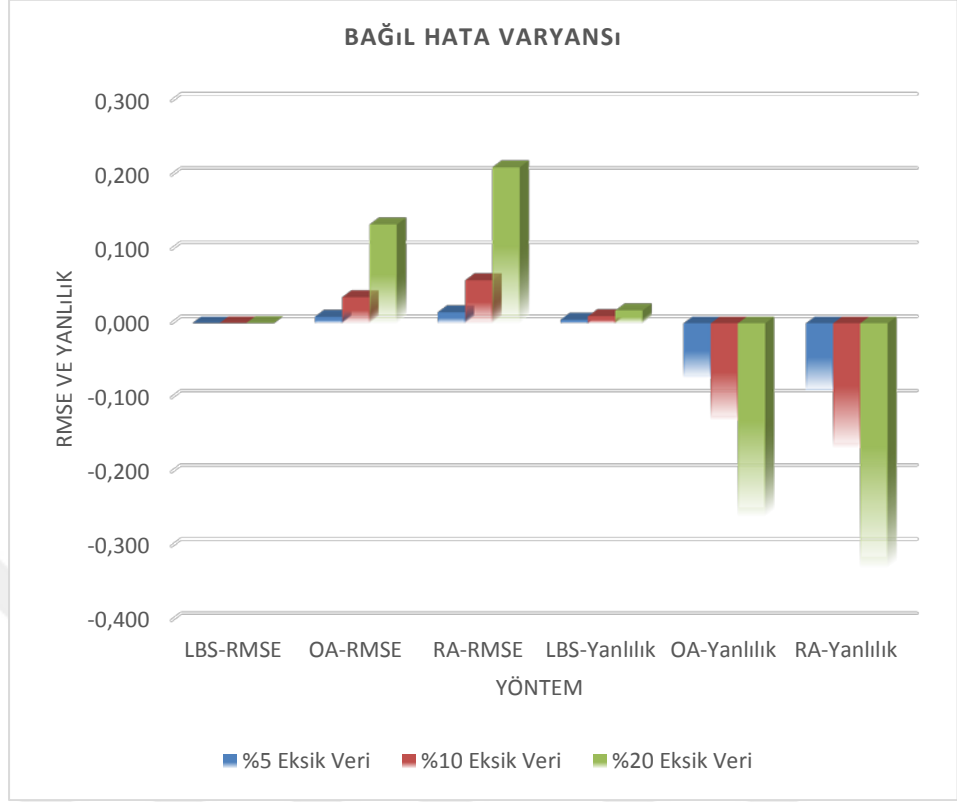


Şekil 2. Evren Puanı Varyansına ait RMSE ve Yanlılık Değerleri

Şekil 2 incelendiğinde elde edilen evren puanı varyansı istatistiğinde en düşük RMSE değerinin LBS yönteminde elde edildiği görülmektedir. Sıfıra en yakın RMSE değerlerinin elde edildiği LBS yönteminin en kararlı yöntem olduğu söylenebilir. OA yöntemi en yüksek RMSE değeri ile en yüksek hataya sahip yöntem olmuştur. Kayıp veri yöntemlerinden bağımsız olarak kayıp veri oranı arttıkça hata değerinin arttığı görülmektedir.

Evren puanı varyansı istatistiğinde en yanlı kestirimler OA yönteminde elde edilmiştir. OA ve RA yönteminde yanlılık değerleri negatif işaretlidir. LBS yönteminde yanlılık değerleri pozitif işaretli ve mutlak değerce en küçüktür. LBS yönteminde yanlılık değerlerinde %5 ve %10 eksik veri oranlarında %10 oranın lehine 0,0003 kadar fark bulunmaktadır. OA ve RA yöntemlerinde kayıp veri oranı arttıkça yanlılık değerlerinin sıfırdan uzaklaştığı görülmektedir.

Bağıl hata varyansı istatistiğine ait grafik Şekil 3 te yer almaktadır.

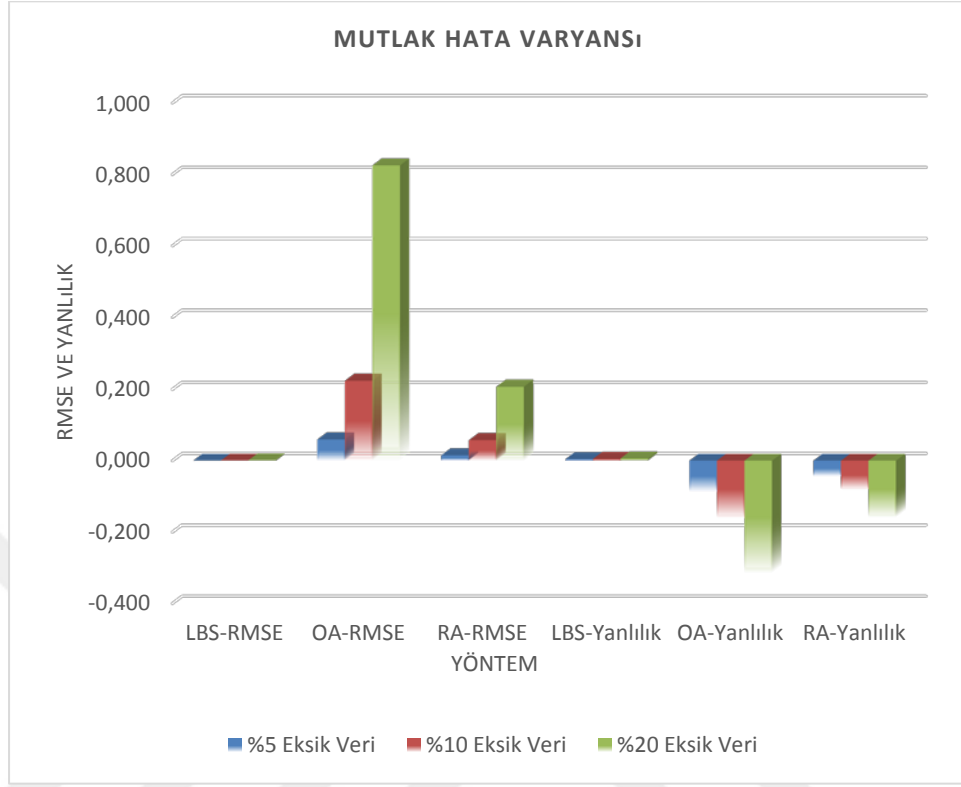


Şekil 3. Bağıl Hata Varyansına ait RMSE ve Yanlılık Değerleri

Şekil 3 incelendiğinde elde edilen bağıl hata varyansına en düşük RMSE değerinin LBS yönteminde elde edildiği görülmektedir. OA ve RA yöntemi kullanıldığında elde edilen hata değerleri birbirine yakına rağmen OA yönteminde daha kararlı kestirimler elde edildiği görülmektedir.

Bağıl hata varyansı istatistiğinde yanlılık değerlerine bakıldığında RA yöntemi kullanılan verilerde yanlılık değerleri negatif işaretlidir ve mutlak değerce en büyüktür. LBS yöntemi kullanılan verilerde bağıl hata varyansı için hesaplanan yanlılık değeri pozitif işaretli ve sifira en yakındır. Kayıp veri oranı koşulu baz alındığında eksik değer yanlılık değeri ile düzgün bir örüntü oluşturmuştur. Tüm yöntemlerde kayıp veri oranı arttıkça yanlılık değeri mutlak değerce artmıştır.

Mutlak hata varyansı istatistiğine ait grafik Şekil 4 te yer almaktadır.

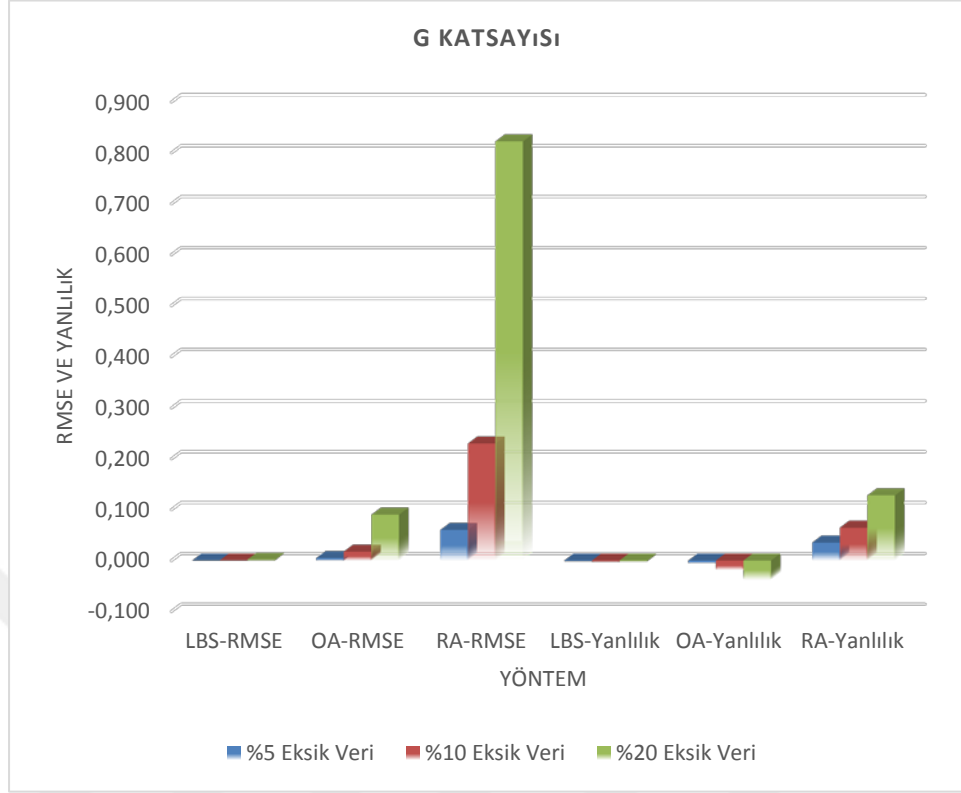


Şekil 4. Mutlak Hata Varyansına Ait RMSE ve Yanlılık Değerleri

Şekil 4 incelendiğinde elde edilen mutlak hata varyansı istatistiğinde en düşük RMSE değerinin LBS yönteminde elde edildiği görülmektedir. En yüksek hataya sahip OA yönteminde RMSE değerleri kayıp veri oranı arttıkça 1 e yaklaşmaktadır. Yöntemlerin tümünde kayıp veri oranı arttıkça hata değerinin arttığı ve daha az kararlı kestirimler elde edildiği görülmektedir.

Mutlak hata varyansına ait yanlılık değerlerine bakıldığında mutlak değerce en büyük değer OA yönteminde eksik veri %20 olduğunda elde edildiği görülmektedir. %10 eksik veri durumunda OA yönteminde hesaplanan yanlılık değeri ikinci sırayı almıştır. LBS yöntemi tüm oranlarda en yansız kestirime sahiptir. Tüm yöntemlerde kayıp veri oranı yanlılık değeri ile düzgün bir örüntü oluşturmuştur. Kayıp veri oranı arttıkça yanlılık değeri artmaktadır.

G katsayısı istatistiğine ait grafik Şekil 5 te yer almaktadır.

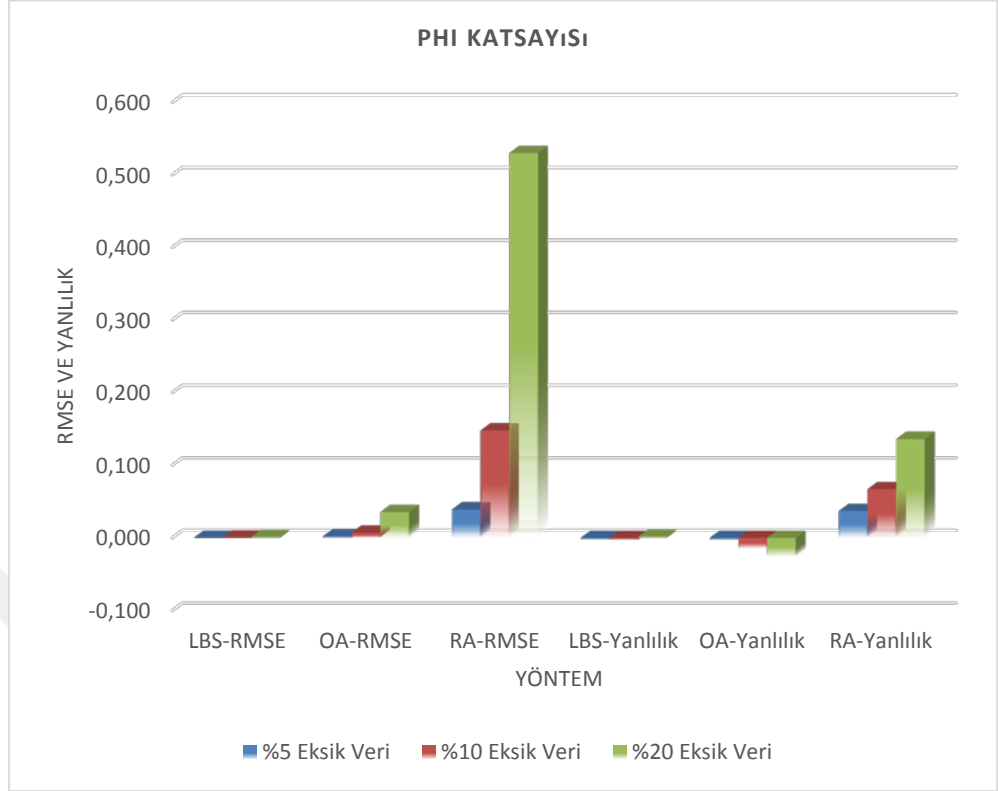


Şekil 5. G Katsayısına ait RMSE ve Yanlılık Değerleri

Şekil 5 incelendiğinde elde edilen G katsayısı istatistiğinde en düşük RMSE değerinin LBS yönteminde elde edildiği görülmektedir. G katsayısı istatistiğinde OA yöntemi ikinci sırada en iyi değeri vermiştir. En yüksek hataya sahip RA yönteminde RMSE değerleri 1'e yakındır.

G katsayısına ait yanlılık değerlerinin genel olarak negatif yönlü olduğu görülmektedir. En yansız kestirim LBS yöntemindedir. Mutlak değerce en büyük yanlılık değeri RA yöntemindedir. OA ve RA yönteminde kayıp veri oranı arttıkça yanlılık değerleri mutlak değerce artmıştır. LBS yönteminde %10'luk eksik veri durumunda 0,0081 gibi küçük bir farka rağmen en yanlı kestirim yapılmıştır.

Phi katsayısı istatistiğine ait grafik Şekil 6 da yer almaktadır.



Şekil 6. Phi Katsayısına ait RMSE ve Yanlılık Değerleri

Şekil 6 incelendiğinde Phi katsayısı istatistiğinde en düşük RMSE değerinin LBS yönteminde elde edildiği görülmektedir. Sıfıra en yakın RMSE değerinin elde edildiği LBS yönteminde kestirimlerin en az hata ile gerçekleştirildiği söylenebilir. En kararlı davranan kayıp veri yöntemi değişmezken RA en yüksek hataya sahip yöntem olmuştur. Kayıp veri yönteminden bağımsız olarak kayıp veri oranı arttıkça hata değerinin arttığı görülmektedir.

Phi katsayısı istatistiğinde de G katsayısına benzer bir durum söz konusudur. Yanlılık değerleri negatif yönlü olma eğiliminde ve en yanlı kestirim RA yönteminde gerçekleşmiştir. OA ve RA yönteminde kayıp veri oranı arttıkça yanlılık artma eğilimi göstermiştir. LBS yönteminde en yansız kestirim bulunan değerler birbirine çok yakın olmasına rağmen sıfıra en yakın olan %20 eksik veri durumunda yaşanmıştır. %10 eksik veri durumu en büyük yanlılık değerine sahiptir.

5. SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırmanın bulgularına ilişkin sonuçlara ve bunlara bağlı olarak önerilere yer verilmiştir.

Bu çalışmada kayıp verilerle baş etmede kullanılan yöntemlerin G kuramı çok değişkenli analizi ile elde edilen istatistikler üzerindeki etkileri farklı oranlardaki kayıp değerler baz alınarak karşılaştırmalı olarak incelenmiştir. Karşılaştırmada RMSE ve yanlılık değerleri ölçüt alınmıştır. Alan yazın incelendiğinde genellikle kayıp veri ile baş etme yöntemlerinin ölçeklerin psikometrik özellikleri bağlamında incelendiği görülmüştür. Kayıp veri yöntemlerinin ölçeklerin güvenilirlik ve geçerliği üzerinde karşılaştırıldığı birçok çalışma bulunmaktadır (Akbaş ve Tavşancıl, 2015; Arıkan ve Soysal, 2018; Başman, 2014; Çokluk ve Kayrı, 2011; Demir, 2013; Doğan ve Soysal, 2017; Enders, 2003; Enders, 2004; Kürşad ve Nartgün, 2015; Raadt, J.Warrens, Bosker ve Kiers, 2019; Shrive F., Stuart, Quan ve Ghali, 2006; Van Ginkel, 2007; Weaver ve Maxwell, 2014; Zhu, 2014). Ayrıca farklılaşan kayıp veri oranı ve örneklem büyüklüğü koşulları altında kayıp verinin G ve Phi katsayısına etkisi araştırılmıştır (Soysal, Karaman ve Doğan, 2018). Bununla birlikte kayıp verilerle baş etmede kullanılan yöntemlerin G kuramı çok değişkenli analizi ile elde edilen istatistikler üzerindeki etkileri açısından incelenmemiştir.

Bu çalışma aynı veri setinden %5, %10 ve %20 oranlarla oluşturulan eksik verili veri setlerinde LBS, OA ve RA yöntemlerinin kullanımıyla elde edilen evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı istatistiklerinin tam veri setinden elde edilen istatistiklerden sapmalarının nasıl değiştiğini incelemiştir.

İlk olarak farklı kayıp veri oranları ile oluşturulan eksik verili veri setlerinde kayıp veri yöntemlerinin evren puanı varyansı, bağıl hata varyansı, mutlak hata varyansı, G ve Phi katsayısı üzerine etkileri incelenmiş ve tüm kayıp veri oranlarında benzer sonuçlar elde edilmiştir. Kayıp veri oranlarının tümünde kayıp veri yöntemi kullanılması sonucu elde edilen istatistikler ile tam veri setinden elde edilen istatistiklerin az da olsa farklılaştığı görülmektedir. Tam verilerden elde edilen istatistikler eksik verilerden elde edilen istatistiklerle karşılaştırıldığında LBS yöntemi tüm kayıp veri oranlarında referans değerlere en yakın sonucu vermiştir. Bu bulgu, Raadt, J.Warrens, Bosker ve Kiers' in (2019) güvenilirlik üzerine yaptıkları

çalışma sonucu ulaşılan bulgu ile benzerlik, Demir (2013), Enders (2003) ve Enders'ın (2004) çalışmaları neticesinde elde ettikleri bulgular ile farklılık göstermektedir. G ve Phi katsayıları için LBS yöntemi kullanılan veri setlerinde tam veri setinden elde edilen değerlere benzer sonuçlar elde edildiği, kayıp veri oranı arttıkça tam veri setinden elde edilen katsayılardan daha düşük değerler bulunduğu görülmüştür. OA yöntemi tüm kayıp veri oranlarında evren puanı varyansı ve mutlak hata varyansı istatistiklerinde farklılaşmanın en fazla olduğu yöntem olmuştur. OA yöntemi kullanılan veri setlerinden elde edilen G ve Phi katsayı değerleri tam veri setinden elde edilen katsayı değerlerinden daha düşük bulunmuş, kayıp veri oranı arttıkça farklılık az da olsa artmıştır. Kayıp veri oranlarının tümünde RA yöntemi kullanılan veri setlerinden elde edilen G ve Phi katsayıları tam veri setinden elde edilenlerden daha yüksek bulunmuştur. Bağlı hata varyansı, G katsayısı ve Phi katsayısı için ise RA yöntemi ana veri matrisinden elde edilen istatistiklerden farklılaşmanın en fazla olduğu yöntem olmuştur. Bu sonuç, Kürşad ve Nartgün 'ün (2015) çalışması ile farklılık göstermektedir.

Kayıp veri oranı arttıkça tam veriden farklılaşmanın arttığı fakat en yakın değerlerin LBS yöntemiyle elde edilmeye devam ettiği görülmüştür. Daha sonra elde edilen analiz sonuçları tam veri setinden ve kayıp veri yöntemleri ile tamamlanmış verilerden elde edilen değerler karşılaştırılarak değerlendirilmiştir. Farklı oranlarda eksiltelen veri setleri ile tam veri setlerinden elde edilen istatistikler arasındaki sapmalar RMSE ve yanlılık ile incelenmiştir. Kayıp veri oranının her koşulu için LBS yöntemi kullanılan veri setlerinde düşük hatalı ve kararlı kestirimlerin yapıldığı görülmüştür. Evren puanı varyansı ve mutlak hata varyansı istatistiklerinde en yüksek hataya sahip kestirim OA yöntemi kullanılan veri setlerinde elde edilmiştir. Çalışmaya konu olan diğer istatistiklerde en fazla hataya sahip kestirimler RA yöntemi kullanılan verilerde elde edilmiştir.

Tüm istatistiklerde ve kayıp veri baş etme yöntemlerinde kayıp veri oranı arttıkça az da olsa hata değerlerinin arttığı gözlenmiştir. Bu bulgu Bayhan (2018) ve Soysal, Karaman ve Doğan'ın (2018) çalışmalarında ulaştıkları bulgularla tutarlıdır.

Yapılan analiz sonucunda LBS yöntemi kayıp veri oranının her koşulu için tüm istatistiklerde en yansız kestirimin yapıldığı yöntem olmuştur. OA kullanılan veri setlerinde elde edilen evren puanı varyansı ve mutlak hata varyansı istatistikleri

kayıp veri oranlarının tüm koşullarında negatif yönlü bulunmuş ve en yanlış kestirim olmuştur. Bağıl hata varyansına ait yanlışlıklar sıfıra çok yakın veya negatif yönlü olarak hesaplanmış ve en yanlış kestirim RA yöntemi kullanılan veri setlerinde elde edilmiştir. G ve Phi katsayısına ait yanlışlık değerlerinde en büyük değer pozitif yönlü olarak RA yönteminde hesaplanmıştır. Kayıp veri oranı koşulu altında incelendiğinde tüm istatistiklerde oran artışının yanlışlık değerlerininin mutlak değerce büyümesine sebep olduğu söylenebilir.

5.1 Uygulayıcılar için öneriler

Bu çalışmada kayıp veri baş etme yöntemlerinin G kuramı çok değişkenli analizi ile elde edilen istatistikler üzerine etkisi incelenmiştir. Yapılan bu çalışma alan yazındaki kayıp veri baş etme yöntemlerini ölçeklerin psikometrik özellikleri bakımından karşılaştırırken G kuramı kullanımının eksikliği göz önünde bulundurularak yapılmıştır. Araştırmada ulaşılan bulgular çerçevesinde çok değişkenli G kuramı analizinde kullanılan veri setlerinde kayıp değer olması durumunda diğer yöntemlere göre kullanım kolaylığı sağlayan LBS yönteminin kullanılması önerilebilir. Kayıp veri oranının düşük olduğu veri setlerinde ise kayıp verilerin RA atama dışındaki diğer yöntemlerle tamamlanması önerilir.

5.2 Araştırmacılar için Öneriler

Sonraki araştırmalar ve araştırmacılar için aşağıdaki öneriler geliştirilmiştir.

1. Çalışmada tek örneklem büyüklüğü üzerinde çalışılmıştır. Örneklem büyüklüğünün etkisini görebilmek için farklı örneklem büyüklükleri kullanılabilir.
2. Çalışmada TROK yapıda veriler kullanılmıştır. ROK ve İEK yapıdaki veri setleriyle analizler tekrarlanabilir.
3. Kayıp veri baş etme yöntemlerinden LBS, OA ve RA yöntemi kullanılmıştır. Kullanılan kayıp veri yöntemleri çeşitlendirilerek mevcut analizler tekrarlanabilir. LBS, OA gibi "geleneksel" yöntemlerin yanında BM, ÇA gibi "modern" yöntemler de tercih edilebilir.
4. Çalışmada LBS yöntemi kullanılan veri setlerinde örneklem büyüklüğü azalmasına rağmen hata ve yanlışlık değerleri daha iyi sonuçlar vermiş, hesaplanan istatistiklerin ortalamaları referans değerlere daha yakın bulunmuştur. Örneklem büyüklüğündeki azalmanın LBS yöntemini nasıl etkilediği üzerine çalışmalar yapılabilir.

KAYNAKÇA

- Akbaş, U. (2014). Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi. Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Akbaş, U. ve Tavşancıl, E. (2015). Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 38-57. doi.org/10.21031/epod.26476.
- Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.
- Alpar, R. (2003). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş*. Nobel Kitabevi.
- Akın Arıkan, Ç. ve Soysal, S. (2018). Güvenirlik katsayılarının kayıp veri atama yöntemlerine göre incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33 (2), 316-336. doi: 10.16986/HUJE.2018036218
- Atılğan, H. (2019). *Genellenebilirlik Kuramı ve Uygulaması*. Ankara: Anı Yayıncılık.
- Atılğan, H. (2013). Sample size for estimation of G and Phi coefficients in generalizability theory. *Eurasian Journal of Educational Research*, 215-227.
- Atılğan, H. (2004). Genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma. Doktora Tezi. Ankara: Hacettepe Üniversitesi.
- Atılğan, H., Kan, A. ve Aydın, B. (2017). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Anı.
- Axelrod, R. (1997). Advancing the art of simulation in the social sciences. *Simulating Social Phenomena*. 21-40.
- Başman, M. (2014). Likert tipi ölçeklerde kayıp verilere değer atamada yapay sınır ağlarının kullanımı. Yüksek Lisans Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Bayhan, A. (2018). Farklı koşullardaki kayıp veri oranının iç tutarlılığa etkisi. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme Klasik Test Teorisi ve Uygulamaları*. Ankara: Pegem Akademi.
- Brennan, R. L. (2001). *Generalizability Theory: Statistics for Social Science and Public Policy*. New York.
- Buuren, S. v., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* , 45(3), 1-67.
- Crocker, L., & Algina, J. (2006). *Introduction to Classical and Modern Test Theory*. Amerika Birleşik Devletleri: Cengage Learning.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *British Journal of statistical Psychology*, 16, 137-163. doi.org/10.1111/j.2044-8317.1963.tb00206.x.
- Cronbach, L. J., Gleser, G. C., Nanda H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability scores and profiles. New York.
- Çokluk, Ö. ve Kayrı, M. (2011). Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenilirliği üzerindeki etkisi. *Kuram ve Uygulamada Eğitim Bilimleri [Educational Sciences: Theory & Practice]*, 11 (1), 289-309.
- Demir, E. (2013). Kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin incelenmesi. Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Karaman, H., Doğan, N. ve Soysal, S., (2017, 14 Mayıs). *Kayıp veri atama yöntemlerinin genellenebilirlik katsayılarına etkisi*. Çalışma 4.Uluslararası Avrasya Eğitim Araştırmaları Kongresinde bildiri, Denizli, Türkiye.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement* , 419-436. doi.org/10.1177/0013164403261050.
- Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, 8(3), 322-337. doi:10.1037/1082-989X.8.3.322.

- Gilbert, N. (1999). Simulation: A new way of doing social science. *American Behavioral Scientist*, 42(10), 1485-1487.
doi.org/10.1177/0002764299042010002.
- Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6(3), 153-160. doi.org/10.1007/BF02289270.
- Şahin Kürşad, M. ve Nartgün, Z. (2015). Kayıp veri sorununun çözümünde kullanılan farklı yöntemlerin ölçeklerin geçerlik ve güvenilirliği bağlamında karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi* , 6(2), 254-267. doi: 10.21031/epod.95917.
- Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin.
- Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis With Missing Data (2nd ed)* New York: Wiley.
- Medley, D. M., & Meitzel, H. E. (1963). Measuring classroom behavior by systematic. N. L. Gage içinde, *Handbook of Research on teaching*. Chicago: Rand McNally.
- Misztal, M. (2012). Imputation of missing data using R package. *Wydawnictwo Uniwersytetu Łódzkiego*, 131-144.
- Moore, C. T. (2016). *Apply generalizability theory with R*. <https://cran.r-project.org/web/packages/gtheory/gtheory.pdf>: CRAN.
- Novo, A. A., & Schafer, J. L. (2013). *Analysis of multivariate normal datasets with missing values*. CRAN.
- Özçelik, D. A. (2016). *Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation: An International Journal of Theory and Practice*, 7 (4), 353-383. doi.org/10.1076/edre.7.4.353.8937.
- Raadt, A. D., J. Warrens, M., Bosker, R. J., & Kiers, H. A. (2019). Kappa coefficients for missing data. *Educational and Psychological Measurement* , 79 (3), 558–576. doi.org/10.1177/0013164418823249.

- Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. (1965). Generalizability of stratified-parallel tests. *Psychometrika*, *30*(1), 39-56. doi:10.1007/BF02289746.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory A Primer*. Sage.
- Shrive, F., Stuart, H., Quan, H., & Ghali, W. (2006). Dealing with missing data in multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology*, *6* (1), 57. doi:10.1186/1471-2288-6-57.
- Soysal, S., Karaman, H. ve Doğan, N. (2018). The effects of sample size and missing data rates on generalizability coefficients. *Eurasian Journal of Educational Research* (75), 179-196. doi: 10.14689/ejer.2018.75.10.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6.th ed). Boston: Pearson.
- Turgut, F. ve Baykul, Y. (2015). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- Van Ginkel, J. R. (2007). Multiple imputation for incomplete test, questionnaire and survey data. Doktora Tezi, Tilburg Üniversitesi.
- Weaver, B., & Maxwell, H. (2014). Exploratory factor analysis and reliability analysis with missing data: A simple method for SPSS users. *The Quantitative Methods for Psychology*, *10* (2) 143-152. doi: 10.20982/tqmp.10.2.p143.
- Zhu, X. (2014). Comparison of four methods for handing missing data in longitudinal data analysis through a simulation study. *Open Journal of Statistics*, *4* (11), 933-944. doi: 10.4236/ojs.2014.411088.

Analizde Kullanılan R Kodları

```
#analysis of the results from the simulation study

#set working directory
setwd("C:/Users/s.gurel/Desktop/hakan_simulation")
library(readxl)
data <- read_excel("ikideneme.xls",
                  sheet = "Sayfa1", col_names = FALSE)
data=as.data.frame(data)
head(data)
names(data)=c("Person",
              paste("ikises.m1.r",c(1:4),sep=""),
              paste("ikises.m2.r",c(1:4),sep=""),
              paste("ikises.m3.r",c(1:4),sep=""),
              paste("ikises.m4.r",c(1:4),sep=""),
              paste("ikises.m5.r",c(1:4),sep=""),
              paste("ikises.m6.r",c(1:4),sep=""),
              paste("uceses.m7.r",c(1:4),sep=""),
              paste("uceses.m8.r",c(1:4),sep=""),
              paste("uceses.m9.r",c(1:4),sep=""),
              paste("uceses.m10.r",c(1:4),sep=""),
              paste("dortses.m11.r",c(1:4),sep=""),
              paste("dortses.m12.r",c(1:4),sep=""),
              paste("ezgibir.m13.r",c(1:4),sep=""),
              paste("ezgibir.m14.r",c(1:4),sep=""),
              paste("ezgibir.m15.r",c(1:4),sep=""),
              paste("ezgibir.m16.r",c(1:4),sep=""),
              paste("ezgiiki.m17.r",c(1:4),sep=""),
              paste("ezgiiki.m18.r",c(1:4),sep=""),
              paste("ezgiiki.m19.r",c(1:4),sep=""),
              paste("ezgiiki.m20.r",c(1:4),sep=""),
              paste("ritimbir.m21.r",c(1:4),sep=""),
              paste("ritimbir.m22.r",c(1:4),sep=""),
              paste("ritimbir.m23.r",c(1:4),sep=""),
              paste("ritimbir.m24.r",c(1:4),sep=""),
              paste("ritimiki.m25.r",c(1:4),sep=""),
              paste("ritimiki.m26.r",c(1:4),sep=""),
              paste("ritimiki.m27.r",c(1:4),sep=""),
              paste("ritimiki.m28.r",c(1:4),sep=""))
require(tidyr)
data2=gather(data,condition,measurement,c(paste("ikises.m1.r",c(1:4),sep=""),
                                           paste("ikises.m2.r",c(1:4),sep=""),
                                           paste("ikises.m3.r",c(1:4),sep=""),
                                           paste("ikises.m4.r",c(1:4),sep=""),
                                           paste("ikises.m5.r",c(1:4),sep=""),
                                           paste("ikises.m6.r",c(1:4),sep=""),
                                           paste("uceses.m7.r",c(1:4),sep=""),
```

```

paste("uceses.m8.r",c(1:4),sep=""),
paste("uceses.m9.r",c(1:4),sep=""),
paste("uceses.m10.r",c(1:4),sep=""),
paste("dortsdes.m11.r",c(1:4),sep=""),
paste("dortsdes.m12.r",c(1:4),sep=""),
paste("ezgibir.m13.r",c(1:4),sep=""),
paste("ezgibir.m14.r",c(1:4),sep=""),
paste("ezgibir.m15.r",c(1:4),sep=""),
paste("ezgibir.m16.r",c(1:4),sep=""),
paste("ezgiiki.m17.r",c(1:4),sep=""),
paste("ezgiiki.m18.r",c(1:4),sep=""),
paste("ezgiiki.m19.r",c(1:4),sep=""),
paste("ezgiiki.m20.r",c(1:4),sep=""),
paste("ritimbir.m21.r",c(1:4),sep=""),
paste("ritimbir.m22.r",c(1:4),sep=""),
paste("ritimbir.m23.r",c(1:4),sep=""),
paste("ritimbir.m24.r",c(1:4),sep=""),
paste("ritimiki.m25.r",c(1:4),sep=""),
paste("ritimiki.m26.r",c(1:4),sep=""),
paste("ritimiki.m27.r",c(1:4),sep=""),
paste("ritimiki.m28.r",c(1:4),sep="")),factor_key = T)
x=unlist(strsplit(as.character(data2$condition),"."))
x.matrix=matrix(x,ncol=3,nrow=77168,byrow=T)
x.matrix=as.data.frame(x.matrix)
names(x.matrix)=c("Domain","Task","Rater")
data3=cbind(data2,x.matrix)
data3=data3[,c(1,4:6,3)]
data4=aggregate(measurement~Domain+Rater+Person,FUN="sum",data3)

require(gtheory)
require(mice)
require(norm)
#Conduct a multivariate G study.
data4$Person=as.factor(data4$Person)
formula <- "measurement ~(1|Person)+
(1|Rater)+
(1|Domain)+
(1|Rater:Domain)+
(1|Person:Domain)+
(1|Rater:Person)"
g=gstudy(data = data4, formula =
formula,colname.objects="Pearson",colname.scores = "measurement")
d=dstudy(g,colname.objects = "Person", data = data4, colname.scores =
"measurement")

g.var.population=g$`components`[,2]
g.percent.population=g$`components`[,3]

```

```

g.n.population=g$`components`[,4]

d.var.population=d$`components`[,2]
d.percent.population=d$`components`[,3]
d.n.population=d$`components`[,4]

generalizability.population=d$generalizability
dependability.population=d$dependability
var.universe.population=d$var.universe
var.error.rel.population=d$var.error.rel
var.error.abs.population=d$var.error.abs

population.data=c(g.var.population,g.percent.population,g.n.population,
                 d.var.population,d.percent.population,d.n.population,
                 generalizability.population,dependability.population,
                 var.universe.population,var.error.rel.population,var.error.abs.population)
population.data=as.data.frame(t(population.data))
population.data2=cbind(population.data,population.data)
population.data3=cbind(population.data2,population.data)
require(splitstackshape)
population.data <- expandRows(population.data3, count=300, count.is.col=FALSE)

names(population.data)=c("g.variance.pd.list","g.variance.rp.list","g.variance.p.list",
                        "g.variance.rd.list","g.variance.d.list","g.variance.r.list","g.variance.resid.list",

                        "g.percent.pd.list","g.percent.rp.list","g.percent.p.list","g.percent.rd.list","g.percent.d.
                        list","g.percent.r.list","g.percent.resid.list",

                        "g.parameter.pd.list","g.parameter.rp.list","g.parameter.p.list","g.parameter.rd.list","g
                        .parameter.d.list","g.parameter.r.list","g.parameter.resid.list",

                        "d.variance.pd.list","d.variance.rp.list","d.variance.p.list","d.variance.rd.list","d.varia
                        nce.d.list","d.variance.r.list","d.variance.resid.list",

                        "d.percent.pd.list","d.percent.rp.list","d.percent.p.list","d.percent.rd.list","d.percent.d.
                        list","d.percent.r.list","d.percent.resid.list",

                        "d.parameter.pd.list","d.parameter.rp.list","d.parameter.p.list","d.parameter.rd.list","d
                        .parameter.d.list","d.parameter.r.list","d.parameter.resid.list",
                        "generalizability.list","dependability.list","var.universe.list",
                        "var.error.rel.list","var.error.ab.list",

                        "g.variance.pd.mean","g.variance.rp.mean","g.variance.p.mean","g.variance.rd.mean
                        ","g.variance.d.mean","g.variance.r.mean","g.variance.resid.mean",

                        "g.percent.pd.mean","g.percent.rp.mean","g.percent.p.mean","g.percent.rd.mean","g.
                        percent.d.mean","g.percent.r.mean","g.percent.resid.mean",

```

```

"g.parameter.pd.mean", "g.parameter.rp.mean", "g.parameter.p.mean", "g.parameter.rd
.mean", "g.parameter.d.mean", "g.parameter.r.mean", "g.parameter.resid.mean",

"d.variance.pd.mean", "d.variance.rp.mean", "d.variance.p.mean", "d.variance.rd.mean
", "d.variance.d.mean", "d.variance.r.mean", "d.variance.resid.mean",

"d.percent.pd.mean", "d.percent.rp.mean", "d.percent.p.mean", "d.percent.rd.mean", "d.
percent.d.mean", "d.percent.r.mean", "d.percent.resid.mean",

"d.parameter.pd.mean", "d.parameter.rp.mean", "d.parameter.p.mean", "d.parameter.rd
.mean", "d.parameter.d.mean", "d.parameter.r.mean", "d.parameter.resid.mean",
      "generalizability.mean", "dependability.mean", "var.universe.mean",
      "var.error.rel.mean", "var.error.ab.mean",

"g.variance.pd.reg", "g.variance.rp.reg", "g.variance.p.reg", "g.variance.rd.reg", "g.varia
nce.d.reg", "g.variance.r.reg", "g.variance.resid.reg",

"g.percent.pd.reg", "g.percent.rp.reg", "g.percent.p.reg", "g.percent.rd.reg", "g.percent.d
.reg", "g.percent.r.reg", "g.percent.resid.reg",

"g.parameter.pd.reg", "g.parameter.rp.reg", "g.parameter.p.reg", "g.parameter.rd.reg", "
g.parameter.d.reg", "g.parameter.r.reg", "g.parameter.resid.reg",

"d.variance.pd.reg", "d.variance.rp.reg", "d.variance.p.reg", "d.variance.rd.reg", "d.varia
nce.d.reg", "d.variance.r.reg", "d.variance.resid.reg",

"d.percent.pd.reg", "d.percent.rp.reg", "d.percent.p.reg", "d.percent.rd.reg", "d.percent.d
.reg", "d.percent.r.reg", "d.percent.resid.reg",

"d.parameter.pd.reg", "d.parameter.rp.reg", "d.parameter.p.reg", "d.parameter.rd.reg", "
d.parameter.d.reg", "d.parameter.r.reg", "d.parameter.resid.reg",
      "generalizability.reg", "dependability.reg", "var.universe.reg",
      "var.error.rel.reg", "var.error.ab.reg")

data2=read.csv(file ="simulation_results 2.csv",header=F)
data3=read.csv(file ="simulation_results 3.csv",header=F)
data4=read.csv(file ="simulation_results 4.csv",header=F)
data=rbind(data2,data3)
data=rbind(data,data4)
#name the dataset
names(data)=c("i", "missing", "g.variance.pd.list", "g.variance.rp.list", "g.variance.p.list",
", "g.variance.rd.list", "g.variance.d.list", "g.variance.r.list", "g.variance.resid.list",

"g.percent.pd.list", "g.percent.rp.list", "g.percent.p.list", "g.percent.rd.list", "g.percent.d.
list", "g.percent.r.list", "g.percent.resid.list",

```

"g.parameter.pd.list", "g.parameter.rp.list", "g.parameter.p.list", "g.parameter.rd.list", "g.parameter.d.list", "g.parameter.r.list", "g.parameter.resid.list",

"d.variance.pd.list", "d.variance.rp.list", "d.variance.p.list", "d.variance.rd.list", "d.variance.d.list", "d.variance.r.list", "d.variance.resid.list",

"d.percent.pd.list", "d.percent.rp.list", "d.percent.p.list", "d.percent.rd.list", "d.percent.d.list", "d.percent.r.list", "d.percent.resid.list",

"d.parameter.pd.list", "d.parameter.rp.list", "d.parameter.p.list", "d.parameter.rd.list", "d.parameter.d.list", "d.parameter.r.list", "d.parameter.resid.list",
"generalizability.list", "dependability.list", "var.universe.list",
"var.error.rel.list", "var.error.ab.list",

"g.variance.pd.mean", "g.variance.rp.mean", "g.variance.p.mean", "g.variance.rd.mean", "g.variance.d.mean", "g.variance.r.mean", "g.variance.resid.mean",

"g.percent.pd.mean", "g.percent.rp.mean", "g.percent.p.mean", "g.percent.rd.mean", "g.percent.d.mean", "g.percent.r.mean", "g.percent.resid.mean",

"g.parameter.pd.mean", "g.parameter.rp.mean", "g.parameter.p.mean", "g.parameter.rd.mean", "g.parameter.d.mean", "g.parameter.r.mean", "g.parameter.resid.mean",

"d.variance.pd.mean", "d.variance.rp.mean", "d.variance.p.mean", "d.variance.rd.mean", "d.variance.d.mean", "d.variance.r.mean", "d.variance.resid.mean",

"d.percent.pd.mean", "d.percent.rp.mean", "d.percent.p.mean", "d.percent.rd.mean", "d.percent.d.mean", "d.percent.r.mean", "d.percent.resid.mean",

"d.parameter.pd.mean", "d.parameter.rp.mean", "d.parameter.p.mean", "d.parameter.rd.mean", "d.parameter.d.mean", "d.parameter.r.mean", "d.parameter.resid.mean",
"generalizability.mean", "dependability.mean", "var.universe.mean",
"var.error.rel.mean", "var.error.ab.mean",

"g.variance.pd.reg", "g.variance.rp.reg", "g.variance.p.reg", "g.variance.rd.reg", "g.variance.d.reg", "g.variance.r.reg", "g.variance.resid.reg",

"g.percent.pd.reg", "g.percent.rp.reg", "g.percent.p.reg", "g.percent.rd.reg", "g.percent.d.reg", "g.percent.r.reg", "g.percent.resid.reg",

"g.parameter.pd.reg", "g.parameter.rp.reg", "g.parameter.p.reg", "g.parameter.rd.reg", "g.parameter.d.reg", "g.parameter.r.reg", "g.parameter.resid.reg",

"d.variance.pd.reg", "d.variance.rp.reg", "d.variance.p.reg", "d.variance.rd.reg", "d.variance.d.reg", "d.variance.r.reg", "d.variance.resid.reg",


```
"d.percent.pd.reg", "d.percent.rp.reg", "d.percent.p.reg", "d.percent.rd.reg", "d.percent.d.reg", "d.percent.r.reg", "d.percent.resid.reg",
```

```
"d.parameter.pd.reg", "d.parameter.rp.reg", "d.parameter.p.reg", "d.parameter.rd.reg", "d.parameter.d.reg", "d.parameter.r.reg", "d.parameter.resid.reg",  
  "generalizability.reg", "dependability.reg", "var.universe.reg",  
  "var.error.rel.reg", "var.error.ab.reg")
```

```
relative.bias.data=(data[,c(3:143)]-population.data)/population.data
```

```
relative.bias.data=cbind(data[,c(1:2)],relative.bias.data)
```

```
names(relative.bias.data)=names(data)
```

```
relative.bias.data$missing=as.factor( relative.bias.data$missing)
```

```
relative.bias.data=as.data.frame(relative.bias.data)
```

```
relative.bias=with(relative.bias.data,aggregate(cbind(g.variance.pd.list,g.variance.rp.list,g.variance.p.list,g.variance.rd.list,g.variance.d.list,g.variance.r.list,g.variance.resid.list,
```

```
g.percent.pd.list,g.percent.rp.list,g.percent.p.list,g.percent.rd.list,g.percent.d.list,g.percent.r.list,g.percent.resid.list,
```

```
g.parameter.pd.list,g.parameter.rp.list,g.parameter.p.list,g.parameter.rd.list,g.parameter.d.list,g.parameter.r.list,g.parameter.resid.list,
```

```
d.variance.pd.list,d.variance.rp.list,d.variance.p.list,d.variance.rd.list,d.variance.d.list,d.variance.r.list,d.variance.resid.list,
```

```
d.percent.pd.list,d.percent.rp.list,d.percent.p.list,d.percent.rd.list,d.percent.d.list,d.percent.r.list,d.percent.resid.list,
```

```
d.parameter.pd.list,d.parameter.rp.list,d.parameter.p.list,d.parameter.rd.list,d.parameter.d.list,d.parameter.r.list,d.parameter.resid.list,
```

```
generalizability.list,dependability.list,var.universe.list,  
  var.error.rel.list,var.error.ab.list,
```

```
g.variance.pd.mean,g.variance.rp.mean,g.variance.p.mean,g.variance.rd.mean,g.variance.d.mean,g.variance.r.mean,g.variance.resid.mean,
```

```
g.percent.pd.mean,g.percent.rp.mean,g.percent.p.mean,g.percent.rd.mean,g.percent.d.mean,g.percent.r.mean,g.percent.resid.mean,
```

```
g.parameter.pd.mean,g.parameter.rp.mean,g.parameter.p.mean,g.parameter.rd.mean,g.parameter.d.mean,g.parameter.r.mean,g.parameter.resid.mean,
```

```
d.variance.pd.mean,d.variance.rp.mean,d.variance.p.mean,d.variance.rd.mean,d.variance.d.mean,d.variance.r.mean,d.variance.resid.mean,
```

d.percent.pd.mean,d.percent.rp.mean,d.percent.p.mean,d.percent.rd.mean,d.percent.d.mean,d.percent.r.mean,d.percent.resid.mean,

d.parameter.pd.mean,d.parameter.rp.mean,d.parameter.p.mean,d.parameter.rd.mean,d.parameter.d.mean,d.parameter.r.mean,d.parameter.resid.mean,

generalizability.mean,dependability.mean,var.universe.mean,
var.error.rel.mean,var.error.ab.mean,

g.variance.pd.reg,g.variance.rp.reg,g.variance.p.reg,g.variance.rd.reg,g.variance.d.reg,g.variance.r.reg,g.variance.resid.reg,

g.percent.pd.reg,g.percent.rp.reg,g.percent.p.reg,g.percent.rd.reg,g.percent.d.reg,g.percent.r.reg,g.percent.resid.reg,

g.parameter.pd.reg,g.parameter.rp.reg,g.parameter.p.reg,g.parameter.rd.reg,g.parameter.d.reg,g.parameter.r.reg,g.parameter.resid.reg,

d.variance.pd.reg,d.variance.rp.reg,d.variance.p.reg,d.variance.rd.reg,d.variance.d.reg,d.variance.r.reg,d.variance.resid.reg,

d.percent.pd.reg,d.percent.rp.reg,d.percent.p.reg,d.percent.rd.reg,d.percent.d.reg,d.percent.r.reg,d.percent.resid.reg,

d.parameter.pd.reg,d.parameter.rp.reg,d.parameter.p.reg,d.parameter.rd.reg,d.parameter.d.reg,d.parameter.r.reg,d.parameter.resid.reg,

generalizability.reg,dependability.reg,var.universe.reg,
var.error.rel.reg,var.error.ab.reg)~
missing,relative.bias.data,mean))

write.csv2(relative.bias,"relative.bias.csv",row.names=F)

```
rmse.data=(data[,c(3:143)]-population.data)^2  
rmse.data=cbind(data[,c(1:2)],rmse.data)  
names(rmse.data)=names(data)  
rmse.data$missing=as.factor(rmse.data$missing)  
rmse.data=as.data.frame(rmse.data)
```

rmse=with(relative.bias.data,aggregate(cbind(g.variance.pd.list,g.variance.rp.list,g.variance.p.list,g.variance.rd.list,g.variance.d.list,g.variance.r.list,g.variance.resid.list,

g.percent.pd.list,g.percent.rp.list,g.percent.p.list,g.percent.rd.list,g.percent.d.list,g.percent.r.list,g.percent.resid.list,

g.parameter.pd.list,g.parameter.rp.list,g.parameter.p.list,g.parameter.rd.list,g.parameter.d.list,g.parameter.r.list,g.parameter.resid.list,

d.variance.pd.list,d.variance.rp.list,d.variance.p.list,d.variance.rd.list,d.variance.d.list,d.variance.r.list,d.variance.resid.list,

d.percent.pd.list,d.percent.rp.list,d.percent.p.list,d.percent.rd.list,d.percent.d.list,d.percent.r.list,d.percent.resid.list,

d.parameter.pd.list,d.parameter.rp.list,d.parameter.p.list,d.parameter.rd.list,d.parameter.d.list,d.parameter.r.list,d.parameter.resid.list,
generalizability.list,dependability.list,var.universe.list,
var.error.rel.list,var.error.ab.list,

g.variance.pd.mean,g.variance.rp.mean,g.variance.p.mean,g.variance.rd.mean,g.variance.d.mean,g.variance.r.mean,g.variance.resid.mean,

g.percent.pd.mean,g.percent.rp.mean,g.percent.p.mean,g.percent.rd.mean,g.percent.d.mean,g.percent.r.mean,g.percent.resid.mean,

g.parameter.pd.mean,g.parameter.rp.mean,g.parameter.p.mean,g.parameter.rd.mean,g.parameter.d.mean,g.parameter.r.mean,g.parameter.resid.mean,

d.variance.pd.mean,d.variance.rp.mean,d.variance.p.mean,d.variance.rd.mean,d.variance.d.mean,d.variance.r.mean,d.variance.resid.mean,

d.percent.pd.mean,d.percent.rp.mean,d.percent.p.mean,d.percent.rd.mean,d.percent.d.mean,d.percent.r.mean,d.percent.resid.mean,

d.parameter.pd.mean,d.parameter.rp.mean,d.parameter.p.mean,d.parameter.rd.mean,d.parameter.d.mean,d.parameter.r.mean,d.parameter.resid.mean,

generalizability.mean,dependability.mean,var.universe.mean,
var.error.rel.mean,var.error.ab.mean,

g.variance.pd.reg,g.variance.rp.reg,g.variance.p.reg,g.variance.rd.reg,g.variance.d.reg,g.variance.r.reg,g.variance.resid.reg,

g.percent.pd.reg,g.percent.rp.reg,g.percent.p.reg,g.percent.rd.reg,g.percent.d.reg,g.percent.r.reg,g.percent.resid.reg,

g.parameter.pd.reg,g.parameter.rp.reg,g.parameter.p.reg,g.parameter.rd.reg,g.parameter.d.reg,g.parameter.r.reg,g.parameter.resid.reg,

d.variance.pd.reg,d.variance.rp.reg,d.variance.p.reg,d.variance.rd.reg,d.variance.d.reg,d.variance.r.reg,d.variance.resid.reg,

d.percent.pd.reg,d.percent.rp.reg,d.percent.p.reg,d.percent.rd.reg,d.percent.d.reg,d.percent.r.reg,d.percent.resid.reg,

d.parameter.pd.reg,d.parameter.rp.reg,d.parameter.p.reg,d.parameter.rd.reg,d.parameter.d.reg,d.parameter.r.reg,d.parameter.resid.reg,
generalizability.reg,dependability.reg,var.universe.reg,
var.error.rel.reg,var.error.ab.reg)~

missing,rmse.data,sum))
rmse=sqrt(rmse)
write.csv2(rmse,"rmse.csv",row.names=F)

means=with(relative.bias.data,aggregate(cbind(g.variance.pd.list,g.variance.rp.list,g.variance.p.list,g.variance.rd.list,g.variance.d.list,g.variance.r.list,g.variance.resid.list,

g.percent.pd.list,g.percent.rp.list,g.percent.p.list,g.percent.rd.list,g.percent.d.list,g.percent.r.list,g.percent.resid.list,

g.parameter.pd.list,g.parameter.rp.list,g.parameter.p.list,g.parameter.rd.list,g.parameter.d.list,g.parameter.r.list,g.parameter.resid.list,

d.variance.pd.list,d.variance.rp.list,d.variance.p.list,d.variance.rd.list,d.variance.d.list,d.variance.r.list,d.variance.resid.list,

d.percent.pd.list,d.percent.rp.list,d.percent.p.list,d.percent.rd.list,d.percent.d.list,d.percent.r.list,d.percent.resid.list,

d.parameter.pd.list,d.parameter.rp.list,d.parameter.p.list,d.parameter.rd.list,d.parameter.d.list,d.parameter.r.list,d.parameter.resid.list,

generalizability.list,dependability.list,var.universe.list,
var.error.rel.list,var.error.ab.list,

g.variance.pd.mean,g.variance.rp.mean,g.variance.p.mean,g.variance.rd.mean,g.variance.d.mean,g.variance.r.mean,g.variance.resid.mean,

g.percent.pd.mean,g.percent.rp.mean,g.percent.p.mean,g.percent.rd.mean,g.percent.d.mean,g.percent.r.mean,g.percent.resid.mean,

g.parameter.pd.mean,g.parameter.rp.mean,g.parameter.p.mean,g.parameter.rd.mean,g.parameter.d.mean,g.parameter.r.mean,g.parameter.resid.mean,

d.variance.pd.mean,d.variance.rp.mean,d.variance.p.mean,d.variance.rd.mean,d.variance.d.mean,d.variance.r.mean,d.variance.resid.mean,

d.percent.pd.mean,d.percent.rp.mean,d.percent.p.mean,d.percent.rd.mean,d.percent.d.mean,d.percent.r.mean,d.percent.resid.mean,

```

d.parameter.pd.mean,d.parameter.rp.mean,d.parameter.p.mean,d.parameter.rd.mean,
d.parameter.d.mean,d.parameter.r.mean,d.parameter.resid.mean,

generalizability.mean,dependability.mean,var.universe.mean,
var.error.rel.mean,var.error.ab.mean,

g.variance.pd.reg,g.variance.rp.reg,g.variance.p.reg,g.variance.rd.reg,g.variance.d.reg
,g.variance.r.reg,g.variance.resid.reg,

g.percent.pd.reg,g.percent.rp.reg,g.percent.p.reg,g.percent.rd.reg,g.percent.d.reg,g.pe
rcent.r.reg,g.percent.resid.reg,

g.parameter.pd.reg,g.parameter.rp.reg,g.parameter.p.reg,g.parameter.rd.reg,g.paramet
er.d.reg,g.parameter.r.reg,g.parameter.resid.reg,

d.variance.pd.reg,d.variance.rp.reg,d.variance.p.reg,d.variance.rd.reg,d.variance.d.reg
,d.variance.r.reg,d.variance.resid.reg,

d.percent.pd.reg,d.percent.rp.reg,d.percent.p.reg,d.percent.rd.reg,d.percent.d.reg,d.pe
rcent.r.reg,d.percent.resid.reg,

d.parameter.pd.reg,d.parameter.rp.reg,d.parameter.p.reg,d.parameter.rd.reg,d.paramet
er.d.reg,d.parameter.r.reg,d.parameter.resid.reg,

generalizability.reg,dependability.reg,var.universe.reg,
var.error.rel.reg,var.error.ab.reg)~
missing,data,mean))

write.csv2(means,"ortalamalar.csv",row.names=F)

write.csv2(data,"simulation.csv",row.names=F)

```