Eurasian Journal of Educational Research 75 (2018) 179-196

Eurasian Journal of Educational Research
www.ejer.com.tr

The Effects of Sample Size and Missing Data Rates on Generalizability Coefficients

Sumeyra SOYSAL[1] , Haydar KARAMAN[2] , Nuri DOGAN[3]

ARTICLE INFO

Article History:
Received: 1 Oct. 2016
Received in revised form: 5 Sep. 2017
Accepted: 4 Apr. 2018
DOI: 10.14689/ejer.2018.75.10

Keywords
Reliability, G coefficient, phi coefficient, zero imputation, MCAR, generalizability theory, matrix of missing data

ABSTRACT

Purpose of the Study: Missing data are a common problem encountered while implementing measurement instruments. Yet the extent to which reliability, validity, average discrimination and difficulty of the test results are affected by the missing data has not been studied much. Since it is inevitable that missing data have an impact on the psychometric properties of measurement instruments, it was considered necessary to investigate this topic. Depending on the identified need, a simulative study was conducted on the effects of missing data on reliability. The reliability estimates were discussed in terms of generalizability theory (G theory). **Research Methods**: Depending on the research questions, complete data sets having different sample sizes (100, 200, 400, 1000) in weak and strong one-dimensional structures under normal distribution were produced. Missing data sets were created by deleting data at different rates (5%, 10%, 20%, 30%) randomly from the complete sets. **Findings and Results:** When the estimates obtained by missing and complete data sets were compared, it was found that G and phi coefficients were significantly affected for the weak one-dimensional design when the missingness was 20% and more. However, for the strong one-dimensional design, those coefficients were negligibly affected even when the missingness was 30%. Moreover, it was also found that the estimates obtained by missing coded incorrect in particularly weak one-dimensional data were lower than the estimates from missing data matrix. Also error statistics of the weak one-dimensional data based on missing coded incorrect were significantly higher than their strong one-dimensional data counterparts, especially at the rates of 20% and 30% missingness. **Implications for Research and Practice** Thus, missing coded incorrect is not suggested to be used as a missing data treatment method in reliability estimations. Instead, generalizability theory, which allows us to conduct analysis with missing data in matrices, might be recommended.

[1] Corresponding author, Hacettepe University, TURKEY, e-mail: sumeyrasoysal@hotmail.com, ORCID: https://orcid.org/ 0000-0002-7304-1722
[2] Karamanoğlu Mehmet Bey University, TURKEY, e-mail: karaman28haydat@gmail.com, ORCID: https://orcid.org/0000-0002-8212-5565
[3] Hacettepe University, TURKEY, e-mail: nuridogan2004@gmail.com, ORCID: https://orcid.org/ 0000-0001-6274-2016

Introduction

One of the problems often encountered in research data collection and analysis is missing data. Missing data can be defined as the difference between the planned set of data and the obtained set of data (Longford, 2005, p. 13). The data set having no missing data is called a complete data matrix, while the data set with missing data is called an incomplete data matrix. The results obtained with statistical methods applied to a complete data matrix can be quite different from the ones obtained with the same statistical methods applied to an incomplete data matrix (Enders, 2010). This is called the missing data problem in statistics. Missing data are an important problem for all branches of science concerned with collecting numerical data. The problem of missing data is more manifest especially in cases where data collection has considerably high costs (Rubin, Witkiewitz, Andre, & Reilly, 2007). Because the missing data in a data matrix is likely to spoil the structure of the matrix, statistical analyses will yield erroneous results and/or biased statistical estimates. Thus, missing (or lacking) values reduce the quality of the data and may risk the reliability of statistical analyses. Methods to eliminate the problem of missing data should be used in order to raise the quality of measurement results containing missing data (Aydilek, 2013; Howell, 2008).

Two elements playing significant roles in the effects of missing data on statistical estimations is the rate of missing data and the design of missing data. It is expected according to research findings that estimation bias increases as the proportion of missing data in the total data increases (Bakis & Goncu, 2015; Cool, 2000). In parallel to the decrease in the ratio of missing data to the total data, the effects on statistical estimations can also be negligible. On the other hand, missing at random (MAR), missing at completely random (MCAR), or not missing at random (NMAR) are categories of missing data (Enders, 2010; Schlomer, Bauman, & Card, 2010). Since the way missing data behaves in terms of distribution is considerably influential in statistical estimations (Enders, 2010; Schafer & Graham, 2002; Zhu, 2014), statistical tests have been developed to determine the distribution of missing data (Little, 1988).

Although studies concerning missing data were started in the early 1900s, they accelerated with such studies as "Inference and Missing Data" by Rubin (1976) and "Statistical Analysis with Missing Data" by Little and Rubin (1987). A review of studies concerning missing data demonstrates that the studies mostly focus on the effects of statistical analysis results on missing data and that they also focus on the effects of value assignment to missing data in different methods on statistical analysis results. A great number of studies conducted by scientists of different branches on the effects of missing data on estimated statistics as well as studies about the effects of missing data assignment methods on statistical analysis results are available in the literature.

For instance, Kose and Oztemur (2014) compared the techniques of multiple coding, listwise deletion and pairwise deletion in their study concerning the variance analysis of missing data and its effects on t test results. Gu and Matloff (2015) also used the same three techniques to study the effect of missing data on regression analysis. They concluded that the method of deletion according to matching had better performance than the other two methods. Cool (2000) investigated the effects of deletion and averaging for missing data on regression estimations and concluded that deletion methods reduced the power of statistical analyses since they shrank sample size. Bakis and Goncu (2015), in their study about biased estimations made by using two coding methods for the incomplete data about flow rate measurements of a stream, concluded that an excessive rate of missing data increased the level of bias in both methods. Ser and Bati (2015), in their study of repeated data in animal husbandry, pointed out that they used a multiple coding method for the missing data in their analysis of general linear mixed model and obtained results similar to the ones obtained through complete data. Shang, Liu, Cheng and Cheng (2016) researched the effects of missing data on the results of component analysis. In a study performed by Yilmaz (2014) on missing data in the field of medicine it was found that coding missing data through closest neighborhood and random forests methods could produce similar solutions in problems of statistical classification, and that the random forests method was preferable in highly related data sets. Some researchers compared the classical techniques with current coding techniques for missing data and they found that multiple coding and maximum likelihood were more advantageous than classical techniques (Allison, 2001; Aydilek, 2013; Baraldi & Enders, 2009; Graham, 2009; Graham, 2012; Nakai, & Ke, 2011; Piggot, 2001; Sari, 2012; Schlomer, Bauman, & Card, 2010). Statistical approaches such as ANOVA, longitudinal development models, structural equation models, regression, logistic regression, general linear models and classification models formed the basis for the comparisons of missing data coding in the abovementioned studies. Allison (2001), Horton and Clainman (2007), Soley-Bori (2013), Whang, Zhang and Tong (2014), on the other hand, considered 9 different software programs and 13 missing data coding techniques in their study introducing the techniques for missing data coding and the statistical software to apply the techniques.

Peng, Harwell, Liou and Ehman (2002), in their study aiming to identify which methods had been used for missing data in articles published in journals of education, examined the studies published in 11 journals in the period between 1998 and 2002. The researchers pointed out that missing data were available in 54% of the studies in the 11 journals, whereas there was no information about data in 18% of the studies. They also found that listwise deletion was used in 87% of the studies, with pairwise deletion used in 7% of the studies, no explanations offered in 3%, and five different methods of coding used in the remaining 3% of the studies.

On reviewing the literature, it was found that the number of studies concerning the effects of missing data on the psychometric properties of measurement instruments used in education and in psychology were limited. In one such study, Weaver and Maxwell (2014) researched the effects of coding missing data on the basis of expectation maximization technique on exploratory factor analysis and on the results of reliability and found it more useful than average coding on the basis of data deletion. Demir (2013) and Cum and Gelbal (2015) researched the effects of missing data coding on confirmatory factor analysis model-data fit values and obtained evidence that relatively new missing data coding methods yielded better results. Nartgun (2015) compared the methods of deletion based on a list, series mean, mean of nearby points, multiple coding and regression coding, which were used in solving the problem of missing data under such conditions as completely random missing mechanism, normal distribution, one-dimensional structure, different sample sizes (n=150; n=650) and different rates of missing data (5%; 10%; 20%). Comparisons were made through the psychometric properties of the scales (eigenvalue, explained variance and Cronbach's alpha) and through statistics calculated from the scores.

Although the literature review showed that the problem of missing data was a common problem encountered in implementing measurement instruments, the review also made it clear that the direct effects on the psychometric properties of measurement instruments were not often considered. In particular, without any methods of missing data imputation and missing data deletion, the issues of how and to what extent reliability and validity of measurement results and such statistics as average discrimination and difficulty are influenced by missing data were not investigated in any depth.

It is common for participants in quantitative studies not to give a response due to various reasons when they are given achievement tests, attitude scales, questionnaires, etc. Participants may sometimes leave a question unanswered due to such reasons as having no idea, failing to find an appropriate answer, skipping a question inadvertently, or not marking the answer correctly. However, as the number of answers to measurement instruments decreases or as missing data increases, the amount of information gathered will decrease and the validity and reliability of measurement results will be expected to fall. It is inevitable that missing data will influence the psychometric properties of measurement instruments used in education and psychology. Therefore, it is believed that the effects of missing data on the psychometric properties of measurement instruments need to be researched. Due to this need, a decision was made to study the effects of missing data on reliability – a psychometric property of measurement instruments. The current study differs from others in this respect.

A second and more important aspect of this study is that it analyzes the effects of the rate of missing data on the generalizability (G) and phi (reliability) coefficients. Brennan (2001) demonstrated that the generalizability and reliability coefficients could be calculated from measurement results having missing data on the basis of generalizability theory with appropriate formulae without deleting a responder from the data. Yet the effects of missing data on the G and phi coefficients were not considered by any researchers. The effects of missing data on the G and phi coefficients represents the main question of this study. In addition, a method most frequently used by researchers encountering the problem of missing data in dual scored data is to regard missing data as incorrect answers and to code them blank. The effects of such an approach on reliability estimations constitutes a second question to which this study seeks answers. Thus, the current study searches for answers to this question: What are the effects of missing data on the reliability of measurement results obtained with one-dimensional measurement instruments? The fundamental research question was considered according to the following subproblems:

1. How is the reliability of measurement results having weak one-dimensional structure obtained with blank coding and incorrect coding influenced by varied sample sizes and by the rates of missing data?

2. How is the reliability of measurement results having strong one-dimensional structure obtained with blank coding and incorrect coding influenced by varied sample sizes and by the rates of missing data?

Method

Research Design

This study has a correlational survey design that aims to determine the presence or degree of co-variance between two or more variables (Karasar, 2004). It is also a simulative study.

Data and Conditions

This study analyzes the effects of missing data on the reliability of one-dimensional measurement results under the condition of varied sample sizes and missing data rates.

Differing recommendations are available in the literature for studies regarding G theory and reliability estimations. Kline (1986) states that sample size should be at least 200 in reliability calculations, while Nunnally and Bernstein (1994) point out that sample size should be at least 300 to reduce the amount of errors stemming from samples. Segal (1994), however, states that sample size of 300 would not be adequate and that it would be small in reliability calculations. Charter (2003), on the other hand, says that sample size of 400 could be adequate. Atilgan (2013) points out that the G and phi coefficients can be estimated in a sufficiently unbiased way if sample size is 50, 100, 200 and 300 in calculating the G coefficients but that the G and phi coefficients will be more precise and stable if sample size is 400. By considering the studies in the literature, sample size was determined as N= 100, 200, 400 and 1000 in this study.

On reviewing the studies concerning the effects of missing data, we found that they were often concerned with differing rates of missing data and structures of missing data. Nartgun (2015) and Kose and Oztemur (2014) conducted their research at a completely random mechanism at the rates of 5%, 10% and 20% missing data. Cheng (2016) analyzed the effects of the presence of 20% missing data at a random missing mechanism. Cum and Gelbal (2015) created data sets containing completely random missing data at the rates of 20% and 30% and not completely random missing data at the rate of 20%. Schafer and Olsen (1998) used a real data set under the condition of MAR and at the rates of 35% and 45%. Shang et al. (2016) used data sets containing missing data at the rates of 10%, 20% and 50% under the conditions of MCAR and MAR. Tabachnick and Fidell (2001) state that missing data at the rate of 5% or above at random do not cause serious problems. Therefore, by considering the studies in the literature, the rates of missing data were determined as 5%, 10%, 20% and 30% in this study.

Depending on sample size, two different types of data sets containing 20 items with normal distribution and representative of strong and weak one-dimensionality were created. Item factor loads were manipulated between 0.50-0.85 in the first type, representing strong one-dimensionality, whereas they were free in the second type, representing weak one-dimensionality. The factor structures of both types of data were analyzed according to the unweighted least squares method; and the factor loads for sample sizes of 100, 200, 400 and 1000 were estimated at the intervals of 0.592-0.824, 0.598-0.808, 0.691-0.820 and 0.765-0.832, respectively, in the strong one-dimensional data, while they were estimated at the intervals of 0.058-0.684, 0.064-0.667, 0.046-0.699 and 0.077-0.677, respectively, for sample sizes of 100, 200, 400 and 1000 in the weak one-dimensional data. The fact that the created data sets had one-dimensional structure was confirmed through analyses by using Factor 10.3 software. Data were deleted from these two complete data sets in an MCAR manner (missing at completely random) at the rates of 5%, 10%, 20% and 30%. The transaction of deleting data was repeated 30 times and finally 30 data sets containing missing data for each sample size were obtained.

Analysis of Research Data

Reliability of measurement results obtained with measurement instruments is calculated in different ways depending on probable sources of error, such as raters, time, test forms, items and task, which may be contained in measurement. This study seeks answers to the research questions through G theory, which enables researchers to assess simultaneously the reliability coefficients obtained in different senses depending on the sources of error.

G theory is based on variance analysis, which ensures that inconsistencies that are present or may be present in observed scores are determined with powerful statistical analysis (Brennan, 2001). G theory divides variability in measurement results into categories according to their sources, and it aims to generalize the observed scores of variable or variables that are the object of measurement into population scores accurately. G theory also removes the traditional difference between validity and reliability to a certain extent. There are relative evaluations and absolute evaluations in education and in psychology, and G theory calculates the generalizability (G) coefficient for relative evaluations and dependability (phi) coefficient for absolute evaluations (Brennan, 2001). This study also examines the change in G and phi coefficients under the condition of sample size and missing data rates.

The G and phi coefficients were calculated in this study from direct missing data matrices and from matrices obtained by using the method of missing data coded incorrect respectively for data sets. Calculations were made manually in Excel because the EduG program was sensitive to missing data and the urGENOVA program could not analyze data containing more than 5% missing data. The calculations made in Excel were performed on the basis of Brennan's example (2001;

p. 227) for pxi design containing missing data and of the analyses for the example. Brennan (2001) employed ANOVA, which uses linear equations obtained by equalizing the expected values of squares averages to estimate variance components in balanced designs. Henderson (1953) recommends two methods for variance and co-variance estimations in unbalanced designs. Brennan (2001) used the method called Henderson Method 1 to calculate the G and phi coefficients in designs containing missing data. Based on this method, a T statistics called sum of squared mean scores is used instead of squares average statistics as in ANOVA. Since the sum of squares is a linear combination of squares average, Brennan (2001) points out that a similar estimation of variance components can be made by equalizing the sum of squares to the expected values. Variance estimations based on T statistics are shown in Table 1,

Table 1

Variance Estimations based on T Statistics for Person x Item Design

Source of Variance	df	T	Sum of Squares
person	n_b-1	$\sum_b \breve{n}_b \bar{X}_b^2$	T(b) - T(μ)
item	n_m -1	$\sum_m \breve{n}_m \bar{X}_m^2$	T(m) - T(μ)
person x item	n_+ - n_b - n_m + 1	$\sum_b \sum_m X_{bm}^2$	T(bm) - T(b) - T(m) + T(μ)
μ	**1**	$n_+ \bar{X}^2$	

Variance estimations of the data created in this study based on T statistics were made and the G and phi coefficients were calculated with the help of these estimations. Absolute and relative error variances were obtained with the following formulas, respectively:

$$\sigma^2(\Delta) = \frac{\sigma^2(m)}{\ddot{n}_m} + \frac{\sigma^2(bm)}{\ddot{n}_m}$$

$$\sigma^2(\delta) = \left(\frac{\sum_b (\bar{X}_b - \bar{X})^2}{n_b - 1}\right) - \sigma^2(\Delta)$$

where $\sigma^2(\delta)$ represents relative error variance, $\sigma^2(\Delta)$ represents absolute error variance. $\ddot{n}_m$ is the harmonic average of n_m . All other calculations used in variance estimations can be found in Brennan (2001; pp. 225-237).

Complete data matrices for each condition were initially created in analyzing the data and the G and phi coefficients were calculated for these complete data matrices. After that, the G and the phi coefficients were calculated separately by the method of missing data incorrect and by missing data design of pxi for each incomplete data set created and averages were found for 30 replications. Finally, the averages found for these two methods of estimation were compared with the results obtained for complete data. The root mean square of errors (RMSE) and bias values of error statistics were then calculated and interpreted.

Results

The findings are presented below according to the research questions.

Research question 1: How is the reliability of measurement results having weak one-dimensional structure obtained with blank coding and zero coding influenced by varied sample sizes and by the rates of missing data?

The G and phi coefficients estimated from matrices obtained from weak one-dimensional complete data matrices by the method of blank coding and incorrect coding (zero coding) are shown in Table 2 below.

Table 2

Averages for the G and Phi Coefficients Estimated from Weak One-Dimensional Data

Sample Size	*Type of Matrix*	*Complete Data*		*5%*		*10%*		*20%*		*30%*	
		G	*Phi*	*G*	*Phi*	*G*	*Phi*	*G*	*Phi*	*G*	*Phi*
N=100	Blank	0.617	0.581	0.601	0.567	0.584	0.551	0.554	0.525	0.519	0.493
	Incorrect			0.577	0.546	0.535	0.508	0.473	0.451	0.405	0.389
N=200	Blank	0.547	0.51	0.531	0.496	0.518	0.484	0.48	0.45	0.439	0.414
	Incorrect			0.506	0.474	0.47	0.442	0.4	0.38	0.339	0.324
N=400	Blank	0.626	0.594	0.613	0.582	0.597	0.568	0.562	0.536	0.522	0.5
	Incorrect			0.591	0.562	0.553	0.527	0.484	0.465	0.42	0.405
N=1000	Blank	0.623	0.592	0.608	0.578	0.592	0.564	0.559	0.534	0.519	0.497
	Incorrect			0.586	0.559	0.564	0.538	0.482	0.464	0.411	0.398

Blank: Missing data coded blank; incorrect: Missing data coded incorrect

According to Table 2, when the rate of missing data is 5%, 10%, 20% and 30% for sample size of 100; the G coefficients estimated from missing data matrices (blank coding) are 0.02, 0.05, 0.10 and 0.16, respectively, and the phi coefficients are estimated lower at the rates of 0.02, 0.05, 0.10 and 0.15, respectively. When the rate of missing data is 5%, 10%, 20% and 30% for sample size of 200; the G coefficients estimated are approximately 0.03, 0.05, 0.12 and 0.20, respectively, and the phi coefficients are approximately 0.03, 0.05, 0.12 and 0.18, respectively, which are low. This is similar to the other samples where the rate of missing data is 5%, 10%, 20% and 30% for sample sizes of 400 and 1000 and the G coefficients are 0.02, 0.05, 0.10 and 0.17, respectively, whereas the phi coefficients are approximately 0.02, 0.05, 0.10 and 0.16, respectively.

As is clear from Table 2, the G coefficients estimated from matrices obtained through incorrect coding for sample sizes of 100 and 200, according to complete data, are approximately 0.06, 0.13, 0.23 and 0.34, respectively, whereas the phi coefficients are approximately 0.06, 0.12, 0.22 and 0.33, respectively, which are low. When the rate of missing data is 5%, 10%, 20% and 30% for sample size of 400, the G coefficients estimated for complete data are approximately 0.06, 0.16, 0.23, and 0.33, while the phi coefficients are approximately 0.06, 0.09, 0.22 and 0.33, which are low. The G and phi coefficients estimated through incorrect coding have been estimated lower than the G and phi coefficients estimated through missing data matrices (blank coding) for all rates of missing data and for all sample sizes.

Bias values and RMSE calculated for the G and phi coefficients from matrices obtained by blank coding and incorrect coding from weak one-dimensional missing data matrices are shown in Table 3.

Table 3

Error Statistics Calculated for Weak One-Dimensional Data

Sample Size	*Type of Matrix*	*Error*	*5%*		*10%*		*20%*		*30%*	
			G	*Phi*	*G*	*Phi*	*G*	*Phi*	*G*	*Phi*
N=100	Blank	RMSE	0.005	0.003	0.020	0.022	0.029	0.030	0.040	0.041
		Bias	0.004	0.003	0.020	0.022	0.028	0.030	0.038	0.041
	Incorrect	RMSE	0.010	0.010	0.021	0.021	0.045	0.046	0.075	0.075
		Bias	0.010	0.010	0.020	0.021	0.045	0.045	0.074	0.075
N=200	Blank	RMSE	0.004	0.003	0.006	0.007	0.015	0.015	0.025	0.026
		Bias	0.003	0.003	0.005	0.007	0.013	0.015	0.023	0.026
	Incorrect	RMSE	0.011	0.011	0.021	0.021	0.046	0.046	0.076	0.075
		Bias	0.010	0.010	0.021	0.021	0.046	0.046	0.075	0.075
N=400	Blank	RMSE	0.004	0.003	0.008	0.007	0.017	0.016	0.029	0.028
		Bias	0.003	0.003	0.007	0.007	0.016	0.016	0.028	0.028
	Incorrect	RMSE	0.011	0.011	0.023	0.024	0.052	0.052	0.085	0.085
		Bias	0.011	0.011	0.023	0.023	0.052	0.052	0.085	0.085
N=1000	Blank	RMSE	0.005	0.003	0.007	0.006	0.015	0.014	0.026	0.024
		Bias	0.004	0.003	0.007	0.006	0.015	0.014	0.025	0.024
	Incorrect	RMSE	0.010	0.010	0.020	0.020	0.045	0.045	0.076	0.076
		Bias	0.010	0.010	0.020	0.020	0.045	0.045	0.076	0.076

Blank: Missing data coded blank; Incorrect: Missing data coded incorrect

According to Table 3, the RMSE and bias error values for the G and phi coefficients estimated from missing data matrices for all conditions of sample size increase in parallel to the increase in the rate of missing data. Also, the situation is similar for the G and phi coefficients estimated from matrices obtained through incorrect coding. It is observed that although error values calculated from both missing data matrices and through incorrect coding for all conditions of the rate of missing data are constant in some cases, as the size of sample increases, the error values decrease at least at minimal levels. Besides, it is also evident on comparing the data sets having and not having incorrect coding, regardless of their sample size, that the RMSE and bias values increase in data sets having incorrect coding.

Research question 2: How is the reliability of measurement results having strong one-dimensional structure obtained with blank coding and zero coding influenced by varied sample sizes and by the rates of missing data?

The G and phi coefficients estimated from matrices that are obtained from strong one-dimensional complete data matrices by the method of blank coding and method of incorrect coding are shown in Table 4 below.

Table 4

Averages for the G and Phi Coefficients Estimated from Strong One-Dimensional Data

Sample Size	*Type of Matrix*	*Complete Data*		*5%*		*10%*		*20%*		*30%*	
		G	*Phi*	*G*	*Phi*	*G*	*Phi*	*G*	*Phi*	*G*	*Phi*
N=100	Blank	0.951	0.951	0.950	0.948	0.947	0.945	0.940	0.938	0.933	0.930
	Incorrect			0.941	0.941	0.931	0.930	0.906	0.906	0.877	0.876
N=200	Blank	0.936	0.936	0.934	0.933	0.931	0.929	0.923	0.921	0.913	0.910
	Incorrect			0.926	0.926	0.915	0.915	0.890	0.890	0.861	0.861
N=400	Blank	0.933	0.933	0.930	0.930	0.926	0.926	0.917	0.917	0.905	0.905
	Incorrect			0.922	0.922	0.910	0.910	0.881	0.881	0.848	0.848
N=1000	Blank	0.942	0.942	0.938	0.939	0.935	0.936	0.927	0.928	0.917	0.918
	Incorrect			0.932	0.932	0.921	0.921	0.897	0.897	0.866	0.866

Blank: Missing data coded blank; Incorrect: Missing data coded incorrect

According to Table 4, when the sample size is 100, the G and phi coefficients obtained from complete data sets are the same as those obtained from matrices obtained by blank coding as data and having 5% and 10% missing data. When the rate of missing data is 20% and 30%, the G and phi coefficients were estimated lower than the complete data set and are 0.01 and 0.02, respectively. These findings are also similar for sample size of 200. Similar G and phi coefficients were estimated for the complete data set with sample sizes of 400 and 1000 and for missing data of 5% and 10%, whereas the coefficients were estimated lower (0.02 and 0.03, respectively) for data sets with 20% and 30% missing data. As is evident from Table 3, equal G and phi coefficients were estimated for all conditions of missing data with sample size of 400. On the other hand, the phi coefficients were estimated higher than the G coefficients for sample size of 1000.

Still according to Table 4, the G coefficients estimated from matrices obtained through incorrect coding for sample sizes of 100, 200 and 1000, according to complete data, are approximately lower at the rates of 0.01, 0.02, 0.05 and 0.08 When the rate of missing data is 5%, 10%, 20% and 30% for sample size of 400, the G and phi coefficients estimated according to complete data are approximately lower at the rates of approximately 0.01, 0.03, 0.06 and 0.09. The G and phi coefficients estimated through incorrect coding are lower than those estimated through matrices of missing data for all rates of missing data and all sample sizes.

Bias values and RMSE calculated for the G and phi coefficients from matrices obtained by blank coding and incorrect coding from strong one-dimensional missing data matrices are shown in Table 5 below.

Table 5

Error Statistics Calculated for Strong One-Dimensional Data

Sample Size	*Type of Matrix*	*Error*	*5%*		*10%*		*20%*		*30%*	
			G	*Phi*	*G*	*Phi*	*G*	*Phi*	*G*	*Phi*
N=100	Blank	RMSE	0.005	0.003	0.020	0.022	0.029	0.030	0.040	0.041
		Bias	0.004	0.003	0.020	0.022	0.028	0.030	0.038	0.041
	Incorrect	RMSE	0.010	0.010	0.021	0.021	0.045	0.046	0.075	0.075
		Bias	0.010	0.010	0.020	0.021	0.045	0.045	0.074	0.075
N=200	Blank	RMSE	0.004	0.003	0.006	0.007	0.015	0.015	0.025	0.026
		Bias	0.003	0.003	0.005	0.007	0.013	0.015	0.023	0.026
	Incorrect	RMSE	0.011	0.011	0.021	0.021	0.046	0.046	0.076	0.075
		Bias	0.010	0.010	0.021	0.021	0.046	0.046	0.075	0.075
N=400	Blank	RMSE	0.004	0.003	0.008	0.007	0.017	0.016	0.029	0.028
		Bias	0.003	0.003	0.007	0.007	0.016	0.016	0.028	0.028
	Incorrect	RMSE	0.011	0.011	0.023	0.024	0.052	0.052	0.085	0.085
		Bias	0.011	0.011	0.023	0.023	0.052	0.052	0.085	0.085
N=1000	Blank	RMSE	0.005	0.003	0.007	0.006	0.015	0.014	0.026	0.024
		Bias	0.004	0.003	0.007	0.006	0.015	0.014	0.025	0.024
	Incorrect	RMSE	0.010	0.010	0.020	0.020	0.045	0.045	0.076	0.076
		Bias	0.010	0.010	0.020	0.020	0.045	0.045	0.076	0.076

Blank: Missing data coded blank; Incorrect: Missing data coded incorrect

An examination of Table 5 makes it clear that the RMSE and bias error values for the G and phi coefficients estimated from missing data matrices (blank coding) for all conditions of sample size increase in parallel to the increase in the rate of missing data. The situation is similar for the RMSE and bias values of the G and phi coefficients estimated from matrices, which are obtained through incorrect coding. It is observed that although error values calculated from both missing data matrices and through incorrect coding for all conditions of the rate of missing data are constant in some cases. The error values decrease at least at minimal levels as the size of sample increases. Moreover, it is also evident on comparing the data sets having and not having incorrect coding regardless of their sample size that the RMSE and bias values increase in data sets having incorrect coding.

Discussion and Conclusion

This study examined the effects of missing data on measurement results and also considered reliability estimates from the aspect of generalizability theory. Studies available in the literature mostly approached the problem of missing data from the aspect of methods for treating missing data and made evaluations by comparing the results for complete data with the ones for treating missing data. However, they did not analyze the psychometric effects of missing data measurement results on statistical analyses without methods for treating missing data.

This study investigated how the rates of missing data in data sets with differing sample sizes and normal distribution influenced the generalizability and phi coefficients when a method for missing data coding was not used.

First, the effects of missing data rates in weak one-dimensional data on G and phi coefficients according to sample sizes were examined, and similar results were obtained for both of these coefficients in weak one-dimensional data. On comparing the estimates made from complete data and the ones made from data with missing data, it was found that the greatest fall was in the data with 20% missing data and especially in the data with 30% missing data. A further conclusion was that the estimates had not been affected greatly by sample sizes. The rate of missing data for a weak one-dimensional set of data having a rate of missing data of 20% and above affected the G and phi coefficients considerably.

Second, the effects of missing data rates and sample sizes in strong one-dimensional tests on G and phi coefficients were investigated, and it was found that the estimates made from missing data were minimally lower than those made from complete data, even in cases with 30% missing data. Thus, it was concluded that sample size did not affect estimates for strong one-dimensional data substantially either.

Estimation errors for the G and phi coefficients obtained from missing data matrices of strong and weak one-dimensional data were analyzed in terms of RMSE and bias statistics. It was found that as the rate of missing data for each condition of sample size increased, error values increased more in weak one-dimensional data and that it increased at minimal levels in strong one-dimensional data. It was also found that RMSE and bias values either did not change or decreased at minimal levels as sample size increased in both weak and strong one-dimensional data for each condition of missing data rates. On evaluating all these conditions together, it was found that error statistics for weak one-dimensional data were bigger than those for strong data.

A method that researchers frequently employ when they encounter missing data in binary data matrices is to regard missing data as incorrect answers and to code them zero. This study also examined the effects of this method and concluded that estimates made by incorrect (zero) coding, especially in weak one-dimensional data, were lower than those made through missing data matrices. On comparing the RMSE and bias values for the G and phi coefficients estimated from missing data matrices with those for matrices obtained by incorrect coding, it was found that the errors based on incorrect coding were higher, which was a remarkable finding. In a similar vein, the error statistics for weak one-dimensional data based on incorrect coding were found to be significantly higher than those for strong one-dimensional data, especially at 20% and 30% rates of missing data. Based on this research finding, it may be said that the incorrect coding method should not be used as a method for treating missing data since reliability estimates with incorrect coding yields biased results. Instead, by considering the fact that the G coefficient obtained in one-faced designs is equal to Cronbach’s alpha, G theory, which enables one to perform analyses with missing data matrices in calculating the reliability of measurement results, is highly recommended.

Another remarkable result obtained in this study was that the G and phi coefficients grew ever closer as sample size increased in strong one-dimensional designs and that the phi coefficient was estimated to be bigger than the G coefficient when the sample size was 1000. Yet the phi coefficient is mathematically smaller than (or equal to) the G coefficient in generalizability analyses for balanced designs. Brennan (2001) states that this situation stems from using different quadratic forms to calculate the T statistics in unbalanced designs.

This study, which aimed to draw the reader’s attention to the fact that the reliability of measurement results could be calculated with G theory, was conducted with binary data. Besides repeating the existing analyses with polytomous data, they can also be performed at differing levels of the conditions in a study. The effects of methods for treating missing data on the reliability of measurement results, which was one of the research problems here, can be analyzed separately in the context of generalizability theory.

References

Allison, P. D. (2001). *Missing data*. Thousands Oaks, CA: Sage Publiation.

Atilgan, H. (2013). Sample size for estimation of G and phi coefficients in generalizability theory. *Eurasian Journal of Educational Research, 51*, 215-227.

Aydilek, İ. B. (2013). *Veri kumelerindeki eksik degerlerin yeni yaklasimlar kullanilarak hesaplanmasi.* (Unpublished doctoral dissertation). Selcuk University, Institute Of Science, Konya.

Bakis, R., & Goncu S. (2015). Akarsu debi olcumlerinde eksik verilerin tamamlanmasi: Zap suyu havzasi ornegi [Completion of missing data in rivers flow measurement: case study of zab river basin]. *Anadolu University*

Journal of Science and Technology A - Applied Sciences and Engineering, *16*(1), 63–79.

Baraldi, A.N., & Enders, C.K. (2009). An introduction to modern missing data analyses. *Journal of School Psychology*, 48, 5-37.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag

Cheng, H. (2016). Principle components analysis with missing values and outliers. Retrieved April 19, 2016, from http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.4.6605&version=3

Cool, A.L. (2000). A review methods for dealing with missing data. Paper presented annual meeting of the Southwest Educational Research Association, Dallas, January 28, TX.

Cum, S., & Gelbal, S. (2015). Kayip veriler yerine yaklasik deger atamada kullanilan farkli yontemlerin model veri uyumu uzerindeki etkisi [The effects of different methods used for value imputation instead of missing values on model data fit statistics]. *Mehmet Akif Ersoy University Journal of Education Faculty*, 35, 87-111.

Demir, E. (2013). Kayip verilerin varliginda coktan secmeli testlerde madde ve test parametrelerinin kestirilmesi: SBS ornegi [item and test parameters estimations for multiple choice tests in the presence of missing data: the case of SBS]. *Journal of Educational Sciences Research, 3* (2), 47-68.

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

Graham, J. W. (2009). Missing data analysis: making it work in the real world. *The Annual Review of Psychology*, *60*, 549-566.

Graham, J. W. (2012). *Missing data: Analysis and design, statistics for social and behavioral sciences*. New York: Springer.

Gu, X., & Matloff, N. (2015). A different approach to the problem of missing data. Retrieved March 20, 2016, from http://arxiv.org/abs/1509.04992

Horton, N. J., & Clainman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician, 61* (*1*), 79-90.

Howell, D. C. (2008). The treatment of missing data. In Outwaite, W. & Turner, S. (Ed.), *The analysis of missing data*. London: Sage Publicaton.

Kose, I. A., & Oztemur, B. (2014). Kayip veri ele alma yontemlerinin t-testi ve ANOVA parametreleri uzerine etkisinin incelenmesi [Examining the effect of missing data handling methods on the parameters of t-test and ANOVA]. *Abant İzzet Baysal University Journal of Faculty of Education*, *14*(*1*), 400-412.

Little, R. J. A., & Rubin D. R. (1987). *Statistical Analysis with Missing* Data. New York: John Wiley and Sons.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(*404*), 1198-1202.

Longford, N.T. (2005). *Missing data and small area estimation*. New York: Springer.

Nakai, M., & Ke, W. (2011). Review of the methods for handling missing data in longitudinal data analysis. *İnternational Journal of Math Analysis*, *5*(*1*), 1-13.

Nartgun, Z. (2015). Comparison of various methods used in solving missing data problems in terms of psychometric features of scales and measurement results under different missing data conditions. *İnternational Online Journal of Educational Sciences*, 7(*4*), 252-265.

Peng, C, J, Harwell, M., Liou, S., & Ehman, L. H., (2002*). Advances in missing data methods and implication for educational research.* Retrieved *March 20, 2016, from www.indiana.edu/~leeehman/missrerfin.pdf*

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

Rubin, D. R. (1976). Inference and missing data. *Biometrika*, *63(3*), 581-592.

Rubin, L. H., Witkiewitz, K., St. Andre, J., & Reilly, S. (2007). Methods for handling missing data in the behavioral neurosciences: don't throw the baby rat out with the bath water. *The Journal of Undergraduate Neuroscience Education (JUNE)*, *5*(*2*), A71-A77.

Sari, İ. K. (2012). *Karma ayristirma analizinde kayip gozlem tahmin yontemlerinin değerlendirilmesi* [Evalution of missing value estimation methods for mixture discriminant analysis]. (Unpublished doctoral dissertation). Selcuk University, Institute Of Science, Konya.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research*, *33*(4), 545-571.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*., *7*(*2*), 147-177.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*(*1*), 1-10.

Ser, G., & Bati, C. T. (2015). Eksik veri analizinde coklu atama yonteminin degerlendirilmesi: Hayvancilikta tekrarli olcum verisi uzerine bir uygulama [Evaluation of multiple ımputation in missing data analysis: an application on repeated measurement data in animal science]. *Turkish Journal of Agriculture - Food Science and Technology*, *3*(*12*), 926-932.

Shang, F., Liu, Y., Cheng, J., & Cheng, H.(2014). Robust principle component analysis with missing data. İn *Proceedings of the 23rd ACM İnternational Conference on*

Conference on İnformation and Knowledge Management, New York, USA, 1149-1158. DOİ=http://dx.doi.org/10.1145/2661829.2662083.

Soley-Bori, M. (2013). Dealing with missing data: Key assumptions and methods for applied analysis. Technical Report no:4. Boston University, School of Public Health, Department of Helath Policy and Managment. Retrived March 20, 2016, from www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf

Tabachnick, B.G., & Fidel L.S. (2001). *Using multivariate statistics (4th ed.).* Needham, Heights, MA: Allyn& Bacon.

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.

Whang, L., Zhang, Z., & Tong, X. (2015). Mediation analysis with missing data through multiple imputation and bootstrap. *Quantitative Psychology Research*, *140*, 341-355.

Yilmaz, H. (2014). *Random Forests yonteminde kayip veri probleminin incelenmesi ve saglik alaninda bir uygulama* [studying the missing data problem in random forests method and an application in health field]. (Unpublished doctoral dissertation). Eskisehir Osmangazi University, Eskisehir.

Zhu, X.P. (2014). Comparison of four methods for handing missing data in longitudinal data analysis through a simulation study. *Open Journal of Statistics*, *4*, 933-944.

Örneklem Büyüklüğünün ve Kayıp Veri Oranının Genellenebilirlik Katsayılarına Etkisi

Atıf:

Soysal, S., Karaman, H., & Dogan, 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

Özet

Problem Durumu: Veri toplama ve bu verilerin analiz edilmesinin temele alındığı araştırmalarda karşılaşılma olasılığı olan en önemli problemlerden biri kayıp veridir. Kayıp veri planlanan veri kümesi ile elde edilen veri kümesi arasındaki fark olarak tanımlanabilir. Alan yazın incelendiğinde kayıp veri problemi ölçme araçlarının uygulanması sonucu karşılaşılan yaygın bir problem olmasına karşın, ölçme araçlarının psikometrik özelliklerine etkisi üzerinde pek durulmamıştır. Özellikle ölçme sonuçlarının güvenirliğinin, geçerliğinin, ortalama ayırıcılık ve güçlük gibi

istatistiklerin kayıp verilerden nasıl ve ne düzeyde etkilendiği konusu pek araştırılmamıştır. Başarı testi, tutum ölçeği, anketler vb. katılımcılara uygulandığı zaman çeşitli sebeplerle bazı katılımcıların cevap vermemesi durumuna çok sık rastlanır. Katılımcılar ölçme araçlarındaki soruları bir fikri olmaması, uygun bir cevap bulamaması, yanlışlıkla soruyu cevaplamadan atlaması veya cevabını doğru bir şekilde işaretlememesi nedenleriyle boş bırakabilmektedir. Ancak ölçme araçlarına gelen cevaplar azaldıkça ya da kayıp veri arttıkça toplanan bilgi azalacak ve ölçme sonuçlarının geçerliği ve güvenirliğinin düşmesi beklenecektir. Kayıp verilerin eğitim ve psikolojide kullanılan ölçme araçlarının psikometrik özelliklerini etkilemesi kaçınılmaz bir durumdur. Dolayısıyla kayıp verilerin ölçme araçlarının psikometrik özellikleri üzerindeki etkisinin araştırılmasına ihtiyaç olduğu düşünülmektedir.

Araştırmanın Amacı: Belirlenen bu ihtiyaca bağlı olarak kayıp verinin ölçme araçlarının psikometrik özelliklerinden güvenirliğe etkisi üzerinde çalışılmasına karar verilmiştir. Bu yönü ile çalışma diğer çalışmalardan farklılık göstermektedir. Çalışmanın ikinci ve daha önemli bir yönü ise kayıp veri oranının genellenebilirlik (G) ve phi (güvenirlik) katsayısına olan etkisini incelemesidir. Brennan (2001), Genellenebilirlik kuramına dayalı olarak kayıp veriye sahip ölçme sonuçlarından uygun formüllerle herhangi bir cevaplayıcıyı verilerden silmeden genellenebilirlik ve güvenilirlik katsayılarının hesaplanacağını göstermiş ancak kayıp verinin G ve Phi katsayısına olan etkisi herhangi bir araştırmacı tarafından incelenmemiştir. Kayıp verilerin G ve Phi katsayısına etkisi bu araştırmanın temel sorusunu oluşturmaktadır. Ayrıca ikili puanlanan verilerde kayıp veri sorunu ile karşılaşan araştırmacıların en sık başvurdukları yöntemlerden biri kayıp verileri yanlış cevap olarak kabul edip sıfır puan ataması yapmaktır. Bu yaklaşımın güvenirlik kestirimine etkisi, bu çalışmayla cevaplamaya çalışılan bir başka sorudur. Dolayısıyla, bu araştırmada normal dağılım altında zayıf ve güçlü tek boyutluluk özelliği gösteren kayıp verili ve sıfır atamayla elde edilen ölçme sonuçlarının güvenirliğinin değişen örneklem büyüklüleri ve kayıp veri oranlarından nasıl etkilendiği sorusuna yanıt aranmıştır.

Araştırmanın Yöntemi: Güvenirlik kestirimleri, hata kaynaklarına bağlı olarak farklı anlamlarda elde edilen güvenirlik katsayılarını aynı anda değerlendirmeyi sağlayan Genellenebilirlik Kuramı açısından ele alınmıştır. Araştırma sorularına bağlı olarak öncelikle normal dağılım gösteren zayıf ve güçlü tek boyutlu yapılarda farklı örneklem büyüklüğüne (N=100, 200, 400, 1000) sahip tam veri setleri üretilmiştir. Bu setlerden tamamıyla seçkisiz olacak şekilde farklı kayıp veri oranlarında (%5, %10, %20,%30) veriler silinerek kayıp verili setler oluşturulmuştur. Araştırma sonuçları tam veri setleri ile kayıp ve sıfır atama yapılmış veri matrislerinden elde edilen G ve phi katsayılarının ortalamaları karşılaştırılarak değerlendirilmiştir. Ayrıca değerlendirmeleri daha isabetli yapabilmek için hata istatistiklerinden hataların kareleri ortalamasının karekökü (RMSE) ve yanlılık (bias) değerleri hesaplanarak yorumlanmıştır.

Araştırmanın Bulguları: Tam veri ile kayıp veri setlerinden elde edilen kestirimler karşılaştırıldığında, zayıf tek boyutlu desenler için kayıp veri oranının %20 ve daha

fazla olduğu durumlarda G ve Phi katsayılarının önemli derecede etkilendiği ancak güçlü tek boyutlu desenler de kayıp veri oranın %30 olduğu durumda dahi bu katsayıların minimal düzeyde etkilendiği bulunmuştur. Örneklem büyüklüğünün her bir koşulu için kayıp veri oranı artıkça hata değerlerinin zayıf tek boyutlu verilerde daha fazla arttığı; güçlü tek boyutlu verilerde ise minimal düzeyde arttığı gözlenmiştir. Kayıp veri oranının her bir koşulu için zayıf ve güçlü tek boyutlu verilerin her ikisinde de örneklem büyüklüğü arttıkça hata ve yanlılık değerlerinin ya değişmediği ya da minimal düzeyde azaldığı görülmüştür. Bütün koşullar bir arada değerlendirildiğinde zayıf tek boyutlu verilere ait hata istatistiklerinin güçlü tek boyutlu verilerden elde edilenlere göre daha büyük olduğu gözlenmiştir. Ayrıca özellikle zayıf tek boyutlu verilerde sıfır atama sonucu elde edilen kestirimlerin kayıp veri matrisinden elde edilen kestirimlerden daha düşük ve sıfır atama yöntemine dayalı olarak zayıf tek boyutlu verilerin hata istatistiklerinin güçlü tek boyutlu verilerin hata istatistiklerinden, özellikle %20 ve %30 kayıp veri oranlarında, önemli derecede yüksek olduğu bulunmuştur.

Araştırmanın Sonuçları ve Önerileri: Dolayısıyla sıfır atama yöntemi ile elde edilen güvenirlik kestirimleri yanlı sonuçlar verdiğinden bu yöntemin güvenirlik kestirimlerinde kayıp veri ile baş etme yöntemi olarak kullanılmaması; bunun yerine ölçme sonuçlarının güvenirliğinin hesaplanmasında kayıp veri matrisleri ile analiz yapmaya olanak sağlayan Genellenebilirlik kuramının kullanılması önerilebilir. Ayrıca kayıp veri matrisleriyle ölçme sonuçlarının güvenirliğinin Genellenebilirlik kuramı ile hesaplanabileceğine dikkat çekmek istenilen bu çalışma iki kategorili veriler ile yürütülmüştür. Mevcut analizler çok kategorili veriler için tekrarlanabileceği gibi araştırmada incelenen koşulların farklı düzeylerinde de gerçekleştirilebilir. Bir başka araştırma problemi olan kayıp veri ile baş etme yöntemlerinin ölçme sonuçlarının güvenirliğine etkisi Genellenebilirlik kuramı bağlamında ayrıca incelenebilir.

Anahtar Kelimeler: Güvenirlik, G katsayısı, phi katsayısı, sıfır atama, MCAR, genellenebilirlik kuramı, kayıp veri matrisi