Investigating the Performance of the Exploratory Graph Analysis When the Data Are Unidimensional and Polytomous

Akif AVCU*

Abstract

The question of how observable variables should be associated with latent structures has been at the center of the area of psychometrics. A recently proposed alternative model to the traditional factor retention methods is called Exploratory Graph Analysis (EGA). This method belongs to the broader family of network psychometrics which assumes that the associations between observed variables are caused by a system in which variables have direct and potentially causal interaction. This method approaches the psychological data in an exploratory manner and enables the visualization of the relationships between variables and allocation of variables to the dimensions in a deterministic manner. In this regard, the aim of this study was set as comparing the EGA with traditional factor retention methods when the data is unidimensional and items are constructed with polytomous response format. For this investigation, simulated data sets were used and three different conditions were manipulated: the sample size (250, 500, 1000 and 3000), the number of items (5, 10, 20) and internal consistency of the scale ($\alpha = 0.7$ and $\alpha = 0.9$). The results revealed that EGA is a robust method especially when used with graphical least absolute shrinkage and selection operator (GLASSO) algorithm and provides better performance in the retention of a true number of dimension than Kaiser's rule and yields comparable results with the other traditional factor retention methods (optimal coordinates, acceleration factor and Horn's parallel analysis) under some conditions. These results were discussed based on the existing literature and some suggestions were given for future studies.

Key Words: Exploratory graph analysis, factor analysis, network psychometrics.

INTRODUCTION

The question of how observable variables should be associated with latent structures have been at the center of the area of psychometrics (Borsboom & Molenaar, 2015). So far, various models were developed to specify this association. However, despite the quantitative increase in numbers and great flexibility of mathematical models used in psychometric studies, the models are surprisingly limited in terms of the paradigm that they are based on.

There are two large families of the models in social sciences to describe the relationships between latent variables and observed variables (Edwards & Bagozzi, 2000). In the first category, the latent traits are considered as the common cause of the observed scores. The model based on such kind of conceptualization is called reflective. Reflective models assume that latent traits cause observed variables (also known as indicators, test items, or symptoms. In reflective models, the indicators are modeled as a function of a common latent variable plus some amount of item-specific error variance. Confirmatory factor analysis (CFA) is one of the most commonly used methods representing reflective models.

Formative models are another broad category to define the relationship between latent structures and observed variables. By this conceptualization, it is accepted that observable variables define the latent structures, not caused by them. The classic example of these kinds of models is the socio-economic status defined by a set of observed variables (e.g. education, job, salary and the district of residency). Principal component analysis (PCA) can be given as a classic example of this kind of model. Using

---

* Research Assistant, Marmara University, Atatürk Faculty of Education, Istanbul-Turkey, avcuakif@gmail.com, ORCID ID: 0000-0003-1977-7592

_____

PCA, data is reduced based on weighted combinations of observed variables to define latent traits (Pearl, 2000).

On the other hand, there is no *"rule of thumb"* when deciding on how many dimensions will retain. In the literature, there are many standard methods for this decision. Kaiser's rule of eigenvalues greater than one rule (KR1: Kaiser, 1960) is the most widely preferred criterion in deciding on how many factors will be retained. This popularity is partly related to its ease of application. However, this method is very sensitive to the number of variables (Gorsuch, 1983) and reliability (Cliff, 1988). Therefore, it may not be effective enough when used in factor retention decisions. An alternative method to KR1 is parallel analysis (PA) developed by Horn (1965). This method is the sample-based adaptation of the KR1 method and has been proposed to alleviate the component indeterminacy problem. Literature shows that this method shows the best performance for component analysis and factor analysis in determining the actual number of factors (Lance, Butts, & Michels, 2006; Velicer, Eaton, & Fava, 2000). Readers are encouraged to look at Kline (2014) for more technical information for KR1, PA and other methods.

More recently, the Acceleration Factor (AF) and Optimal Coordinates (OC) methods were proposed by Raiche, Riopel and Blais (2006) and Raiche (2010). These methods provide non-graphical solutions to Cattell's scree test (1966) to overcome its subjective weakness. AF shows where the elbow of the slope is on the graph and corresponds to the curve's acceleration, i.e. the second derivative. That is, it aims to determine the point where the slope changes abruptly. OC is the other method based on measuring gradients associated with the eigenvalues and preceding eigenvalues to determine the slope's location. It has been stated that AF and OC methods perform better than the KR1 method and approach the performance of PA under certain conditions. (Ruscio and Roche, 2012).

A recently proposed alternative model to traditional reflective and formative approaches is called network modeling. In this approach, there is an assumption that the associations between observed variables are caused by a system in which variables have direct and potentially causal interaction with each other (Eaton, 2015). The usage of network models has provided considerable benefit for understanding complex systems in many different disciplines (Barabási & Pósfai, 2016). In the social sciences, the application of network analysis was adopted firstly to investigate social network structures (eg. Cartwright and Harary, 1956). However, in the following decades, it has been used as an alternative to latent variable modeling in studies to analyze network models of psychological behaviors in an exploratory manner (Borsboom & Cramer, 2013; Schmittmann et al., 2013). After this shift in the application of network modeling, the popularity of the network approach increased and it started to be used intensively in psychology and led to the emergence of a new branch of psychology aimed at predicting network structures in psychological data. This new branch is called network psychometrics (Epskamp, Maris, Waldorp, & Borsboom, 2015).

As with other network models, a psychometric network model consists of a series of nodes (or vertices), a set of connections or links between the nodes (also known as edges) and information regarding the structure of nodes and edges (De Nooy, Mrvar & Batagelj, 2011). In this framework, the nodes represent the psychological indicator variables (e.g. symptoms, behaviors, or faces of latent variables). Traditionally, they are represented by circles in the network structure. On the other hand, the edges represent the node's associations and represented in a network models by lines connecting the nodes.

A more recent paper (Golino & Epskamp, 2017) introduced an innovative way to investigate the dimensionality of psychological constructs by network modeling. This new method is called the EGA. As its name implies, this model is not based on prior assumptions when investigating the dimensionality of a construct. Instead, it approaches the psychological data in an exploratory way. A fascinating feature of EGA is that it enables the visualization of the relationships between variables and allocating variables to the dimensions in a deterministic manner (Golino et al., 2020). For this reason, it is an ideal method to test or reevaluate the theoretical structure of psychological constructs.

In an EGA model, traditionally green (or blue) lines on the network represent positive partial correlations, and red lines correspond to negative partial correlations. In addition, the thickness of the

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology   2

lines gives information about the amount of the correlation as the thicker lines indicate that the partial correlation values approach 1. If the partial correlation values are exactly 0, no line is drawn between the two nodes which implies that the two variables are independent when other variables in the network are conditionally controlled (Pearl, 2000). In figure 1, an exemplary graph of EGA was presented.

Like other psychometric network models, the EGA is also based on Gaussian Graphical Modeling (GGM), which was proposed by Lauritzen (1996). This model estimates the joint distribution of random variables by modeling the inverse of the variance-covariance matrix (Epskamp, Borsboom & Fried, 2018). In this type of modeling, each edge value represents the relationship between a node pair after conditioned to other variables in the model (Epskamp & Fried, 2016). In more concrete terms, partial correlations are used for the construction of networks in the models. If no edges were drawn between nodes, it implies that zero value for partial correlations is estimated. That is, the nodes are not connected in the model and show conditional independence.

Like other statistical methods that use sample data to estimate parameters, correlation and partial correlation values are also affected by sampling variation. Hence, the exact zero values in matrices are rarely be observed in real data. As a result, the estimated networks based on partial correlations become fully connected. Small weights on many edges could possibly reflect weak and potentially spurious partial correlations in this kind of network. These spurious relationships cause a threat to the clear interpretation of networks and replicability. Frequently, a statistical method is used to remove these spurious connections and control network complexity. For estimations based on partial correlations, a commonly used procedure is to apply the least absolute shrinkage and selection operator (LASSO) proposed by Friedman, Hastie and Tibshirani (2008). Because the LASSO can control spurious connections, this method can provide high precision estimates when combined with the community detection algorithm, such as the walktrap algorithm (Pons & Latapy, 2005).

LASSO uses a tuning parameter to remove spurious connections in the model by filtering the network with penalization approach to the inverse covariance matrix. In this way, partial correlation values smaller than a threshold are estimated as exactly zero. The tuning parameter was selected based on minimizing Extended Bayesian information criterion (EBIC) proposed by Chen and Chen (2008). It enables the researcher to control the sparsity of networks (Foygel & Drton, 2010). LASSO is an important part of network modeling because it determines the eventual network structure. It also enables obtaining parsimonious and more interpretable models. In EGA models, a graphical extension of LASSO is used and referred to as GLASSO. In addition, as an alternative to GLASSO, Triangulated Maximally Filtered Graph (TMFG) was proposed. This approach builds a triangulation that enables a score function to maximize. In this way, the data becomes organized in a meaningful structure and modeling becomes possible. The detailed explanations and formulations could be found in Massara, Di Matteo and Aste (2016).

As cited above, the EGA was firstly proposed by Golino & Epskamp (2017). In this paper, they compared the performance of the EGA with five different traditional factor retention methods. These methods are as follows: (a)very simple structure (VSS; Revelle & Rocklin, 1979); (b) minimum average partial procedure (MAP; Velicer, 1976); (c) fit of a different number of factors, from 1 to 10, via BIC and via EBIC; (d) Horn's Parallel Analysis (PA; Horn, 1965); (e) Kaiser-Guttman eigenvalue greater than one rule (Guttman, 1954); (f) EGA.

In the study, these methods were compared with each other by using simulated data sets across different conditions: the sample size (100, 500, 1000 and 5000), the number of factors (2 and 4), the number of items in each factor (5 and 10) and the correlation between the dimensions (.2, .5 and .7). The datasets were generated in two and four dimension structures and as having dichotomous items. The effectiveness of the methods was tested with their estimation rate of a true number of factors. These methods were compared in terms of their performance to extract the true number of dimensions. According to the findings, it was reported that EGA performed better than the traditional factor retention methods especially when the datasets were simulated as having four dimensions and when the number of items in each dimension was five. It was also stated that EGA was found to be the only method giving satisfactory results in all conditions. All in all, this study confirmed the superiority of

EGA to other traditional methods under some conditions. As this study revealed, EGA is suitable to be used with multidimensional datasets.

On the other hand, the reason why multidimensional datasets were preferred in this recent study is that EGA framework was available to be used only with multidimensional datasets, but a recent revision allowed the examination of unidimensional datasets. In this way, practical limitations to test the effectiveness of EGA with unidimensional datasets were eliminated. There are a number of important reasons to examine unidimensionality in tests. First of all, there is a need to calculate the α coefficient for the overall test (Dunn, Baguley, & Brunsden, 2014). In addition, unidimensionality indicates the presence of a common underlying cause or a coherent set of homogeneous causes (DeVellis, 2017). Based on these facts, Golino & Epskamp (2017) recommended testing the performance of EGA with unidimensional datasets composed of polytomously scored items.

Considering the richness of outputs (such as centrality measures, node strength measures, item stability statistics and entropy fit index) EGA provide to evaluate psychometrical properties of scales (Golino & Christensen, 2020), it is assumed that test developers will use EGA with increasing frequency in the future. In addition, some psychological traits like depression (Beard et al. 2016), anxiety (Fisher et al., 2017) or addiction are measured based on the symptoms they are relied on. DiFranza and his colleagues (2002) suggested considering these symptoms as interconnecting networks rather than indicators caused by latent traits. It is assumed that such kinds of understanding of psychopathological symptoms can contribute more to our understanding of disorders (Beard et al. 2016). For this reason, it is fair to assume that use of EGA will increase in the future.

Purpose of the Study

In this regard, the aim of this study was set as the comparison of the performance of EGA with traditional factor retention methods when the data is unidimensional and items are scored in polytomous response format.

METHOD

Data Simulation Procedure

In the current study, three different conditions were manipulated: the sample size (250, 500, 1000 and 3000), the number of items (5, 10, 20) and the internal consistency level (α = 0.7 and α = 0.9). The conditions of the study were determined by taking into account the features of the scales in the existing psychology literature. Related literature shows that the number of items in unidimensional measurement tools show variance. For example, the Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin (1985) consists of five items while the Center for Epidemiologic Studies Depression Scale (Radloff, 1977) consists of twenty items. For this reason, a number of items in simulated data sets were allowed to vary between these observed values (5,10,20). In addition, in order to consider a test to be reliable, the lower threshold value was proposed as .7 (Nunnaly 1978). On the other hand, if the α level is above .90, it is regarded as the test has a good level of α. Accordingly, the data sets were simulated as half of them had α at lower threshold (α = 0.7) while another half of the datasets were simulated as having α level regarded as good (α = 0.9). Finally, the sample size of n=250 is generally regarded as the minimum number when applying factor retention methods (Cattell, 1978). For this reason, the simulated datasets were arranged to had a sample size of at least 250 while n=500, n=1000 and n=3000 conditions were also selected when generating data sets. Based on these facts, 24 different conditions were created with a 4x3x2 design. Finally, in line with the main aim of this study, all of the data sets were simulated as having unidimensional structure and datasets were generated as if the items were scored between 1-5 intervals.

For each condition, data simulation was repeated 100 times to obtain more stable results. This process resulted in generating 2400 datasets. The reported results in this study reflect the arithmetic average

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

4

of the iterations. The data simulation was performed with *mirt* package (Chalmers, 2012) in R program (R core team, 2019).

Analysis Procedure

EGA analyses were carried out using the *EGAnet* package available in R statistical environment (Golino & Christensen, 2020). The tuning parameter for GLASSO was determined based on EBIC to obtain a sparser network. In this study, this parameter was set at 0.5, which is a default option in *EGAnet*. On the other hand, the *nFactors* package (Raiche, 2010) was used for applying OC, AF, PA and KR1 factor retention methods.

The assessment of how accurate the correct number of dimensions is extracted was made based on extraction accuracy index and bias indices, as Garrido, Abad & Posada (2016). Factor extraction accuracy index was calculated at two stages: (1) coding correct estimation of the true number of factors as 1 and incorrect estimation of castors as 0, (2) taking the arithmetic mean of coded scores. For instance, when 100 datasets were analyzed, if the true number of factors extracted for 50 datasets, the accuracy index was computed as 0.5. On the other hand, the bias index was calculated as a subtraction of the estimated number of dimensions from a true number of dimensions. For instance, for a unidimensional dataset if the estimated number of the dataset is 1, the bias index is calculated as 0 while if the estimated number is 2, the bias value becomes 1. Therefore, a bias value of 0 indicates the correct number of dimensions are extracted perfectly while a bias values far from 0 indicates the poor performance of the corresponding method. Similar to the accuracy index, the values of bias in the results section represents the arithmetic mean of 100 iterations.

RESULTS

The average accuracy index values and corresponding standard deviations obtained from 100 iterations were given in Table 1. When the sample size was set as 250 and datasets contained five items, all of the methods estimated the correct number of factors perfectly regardless of the α level. As the number of items was increased to ten and α level was 0.7, EGA (LASSO) could extract unidimensional structure for 79% while this rate was 49% for EGA(TMFG). Both algorithms of EGA method outperformed the traditional KR1 method. When the α level has risen to 0.9, EGA (LASSO) method estimated the correct number of dimensions for 99% of datasets, whereas EGA (TMFG) method's percentage drops to 9%. On the other hand, for the other four traditional methods, the average accuracy rates were 100%. In particular, EGA (LASSO) method yielded comparable results with traditional methods when the alpha level was 0.9. Finally, for data sets containing twenty items, the accuracy rate of EGA(LASSO) was 2% and 52% for the conditions where the α was 0.7 and 0.9 respectively, whereas accuracy rates of EGA (TMFG) were 0% for both α levels. The only method EGA(LASSO) outperformed was KR1 while EGA (TMFG) yielded the worst accuracy rates.

For the datasets with n=500 sample size condition, all of the methods examined were perfectly estimated unidimensional structure when the sample size contained five items. This result didn't show a difference across α levels. On the other hand, when the number of items was increased to 10 and α level was 0.7, the average accuracy rate of EGA(LASSO) and EGA(TMFG) was found to be 0.99 and 0.45 respectively. EGA(LASSO) outperformed the traditional KR1 method while EGA(TMFG) method yielded the lowest accuracy levels. As the α level increased to 0.90, EGA(TMFG) was the only method that provided an imperfect accuracy rate (%22). Finally, as the number of items in the datasets was increased to 20, only AF performed a perfectly estimated true number of dimensions when the α was set to be 0.7 while AF and PA performed perfectly when the α level was 0.90. On the other hand, EGA methods yielded the worst accuracy rates.

For the n=1000 sample size condition, when the dataset contained five items, all of the methods extracted the correct number of dimensions perfectly while imperfect rates were obtained for EGA(TMFG) with accuracy rates of 0.59 and 0.26 depending on the α level for the datasets contained ten items. Finally, as the number of items was set to be 20, the EGA(LASSO) method's accuracy rates

were 68% and 99% for the α levels of 0.7 and 0.9, respectively. On the other hand, EGA(TMFG) yielded perfectly inaccurate results.

For datasets where the sample size was 3000, the accuracy rate for EGA (LASSO) was 99% when α was 0.7 and the number of items was 20, while it was 100% in other conditions. For EGA (TMFG) method, the accuracy rates for datasets with 10 and twenty items fell to 77% and 0% when the alpha was α = 0.7, while the accuracy rates for the data sets with ten and twenty items and with α value of .9, accuracy rates decreased to 36% and 0% respectively. For the KR1 method, the accuracy rate was 3% for datasets where α = 0.7 and the number of items was 2. For OC, AF and AP methods, a 100% accuracy rate was achieved under all conditions. Lastly, EGA(LASSO) yielded a 99% accuracy rate when α level was 0.7 and datasets contained twenty items while it perfectly estimated true number of dimensions for the rest of the conditions. On the other hand, EGA(TMFG) yielded the lowest accuracy rates when the number of items was 10 and 20. Especially, OC, AF and AP methods yielded perfect accuracy rates under all conditions examined. As could be inferred, based on the number of items, EGA's relative performance against traditional factor retention methods changed dramatically. In addition, for most of the conditions, GLASSO algorithm was superior to TMFG algorithm.

Table 1. Mean Accuracy of Factor Retention Methods

	EGA(LASSO)		EGA(TMFG)		OC		AF		PA		KR1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
n=250												
α = 0.70												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	0.79	0.41	0.49	0.50	0.91	0.29	1.00	0.00	0.91	0.29	0.02	0.14
20 items	0.02	0.14	0.00	0.00	0.58	0.50	1.00	0.00	0.58	0.50	0.00	0.00
α = 0.90												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	0.99	0.10	0.09	0.29	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
20 items	0.52	0.50	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.03	0.17
n=500												
α = 0.70												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	0.99	0.10	0.45	0.50	1.00	0.00	1.00	0.00	1.00	0.00	0.75	0.44
20 items	0.45	0.50	0.00	0.00	0.87	0.34	1.00	0.00	0.86	0.35	0.00	0.00
α = 0.90												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	1.00	0.00	0.22	0.42	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
20 items	0.93	0.26	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.73	0.45
n=1000												
α = 0.70												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	1.00	0.00	0.59	0.49	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
20 items	0.68	0.47	0.00	0.00	0.99	0.10	1.00	0.00	0.99	0.10	0.00	0.00
α = 0.90												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	1.00	0.00	0.26	0.44	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
20 items	0.99	0.10	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
n=3000												
α = 0.70												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	1.00	0.00	0.77	0.42	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
20 items	0.99	0.10	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.03	0.17
α = 0.90												
5 items	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
10 items	1.00	0.00	0.36	0.48	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
20 items	1.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

6

The calculated bias values for the factor retention methods across conditions were given in Table 2. If the datasets contained five items, EGA(LASSO) provided unbiased estimates of the correct number of dimensions. As the number of items in the datasets was increased to 10 and the sample size of n=250, the bias value was estimated to be 0.33 0.01 for α levels of 0.7 and 0.9, respectively. As the sample size of datasets was increased to 500, EGA(LASSO) yielded 0.01 and 0 bias for α levels of 0.7 and 0.9. When the sample size was n=1000 and n=3000, EGA(LASSO) yielded no bias when the item number was 10. For the datasets containing twenty items, if the sample size was n=250, the bias value was 2.41 for α level of 0.7 and 1.39 for α level of 0.90. On the other than, the bias value of 1.39 has very large standard deviation value which indicated that, there was a variation across the datasets in terms of the bias value calculated. As the sample size was increased to 500, 1000 and 3000, the bias values calculated showed a decrease compared to n=250 condition. Similar changes were also observed for EGA(TMFG) across the conditions while EGA(TMFG) performed worse than EGA(LASSO) in general. On the other hand, other traditional estimation methods provided almost perfect results especially when the sample size was n=1000 and n=3000.

Table 2. Mean Bias Error of Factor Retention Methods

	EGA(LASSO)		EGA(TMFG)		OC		AF		PA		KR1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
n=250												
α = 0.70												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.33	1.02	0.57	0.61	0.10	0.33	0.00	0.00	0.10	0.33	1.42	0.54
20 items	2.41	1.16	2.27	0.75	0.56	0.74	0.00	0.00	0.60	0.84	6.09	0.71
α = 0.90												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.01	0.10	1.00	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20 items	1.39	3.71	2.20	0.75	0.00	0.00	0.00	0.00	0.00	0.00	1.73	0.72
n=500												
α = 0.70												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.01	0.10	0.57	0.54	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.44
20 items	1.37	2.79	2.35	0.69	0.15	0.44	0.00	0.00	0.17	0.47	5.29	0.67
α = 0.90												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.00	0.00	0.85	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20 items	0.07	0.26	2.25	0.74	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.50
n=1000												
α = 0.70												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.00	0.00	0.42	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20 items	0.85	2.74	2.28	0.74	0.01	0.10	0.00	0.00	0.01	0.10	4.50	0.64
α = 0.90												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.00	0.00	0.85	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20 items	0.02	0.20	2.23	0.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
n=3000												
α = 0.70												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.00	0.00	0.25	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20 items	0.01	0.10	2.05	0.73	0.00	0.00	0.00	0.00	0.00	0.00	1.63	0.68
α = 0.90												
5 items	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 items	0.00	0.00	0.70	0.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20 items	0.00	0.00	2.10	0.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

After calculating the accuracy rates and the bias values, a series of factorial ANOVA was performed to examine the effects of conditions altered for each factor retention method. For this analysis, the raw estimated dimension number value was used as the dependent variable. Only eta square ($\eta 2$) effect size values and the significance levels of ANOVA analysis were reported. The significance levels, ** sign denotes significance at _p<0.01_ level and * implies significance at _p<0.05_. The $\eta 2$ values show the magnitudes of the differences between the conditions for each method under investigation. According to Cohen (1988), $\eta 2$ values of 0.14 and above can be regarded as a "_large_" effect size. On the other hand, the effect size for AF method cannot be compared because this method perfectly estimated the true number of dimensions for all 2400 datasets.

For the rest of the methods, it was found that the unique effects of the conditions examined for EGA (GLASSO) method or their two-way and three-way interactions did not have a large effect size. Similar results were observed for OC and PA methods. On the other hand, the item number condition had a large effect size for EGA(TMFG) method. Finally, for the KR1 method, large amounts of the effect size values were observed for each of the conditions examined and their two-way and three-way interactions were found as significant.

Table 3. Effect Sizes of Factorial ANOVA

	EGA(GLASSO)	EGA(TMFG)	OC	AF	PA	KR1
Sample Size (SS)	0.05**	0.01**	0.05**	-	0.05**	0.62**
Number of Items (NI)	0.09**	0.75**	0.04**	-	0.04**	0.91**
Reliability (r)	0.02**	0.01**	0.03**	-	0.03**	0.80**
SS X NI	0.07**	0.01**	0.06**	-	0.06**	0.65**
SS X r	0.01**	0.01	0.05**	-	0.05**	0.36**
NI X r	0.02**	0.03**	0.04**	-	0.04**	0.86**
SS X NI X r	0.01**	0.01	0.06**	-	0.06*	0.45**

**_p<0.01_

DISCUSSION and CONCLUSION

The current study aimed to compare the effectiveness of EGA in extracting the true number of dimensions with traditional methods when the data was unidimensional and composed of polytomous items. This aim was determined based on Golino and Epskamp's (2017) recommendations and literature review showed that no study was conducted so far considering this recommendation. Unlike this study, in the current study, OC and AF methods were included for comparison because these methods are also relatively new compared to more traditional methods like PA and KR1 and their inclusion on relatively new methods is believed to increase existing knowledge on the effectiveness of EGA.

As a result of this study, it has been observed that EGA (LASSO) successfully extracted unidimensional structure perfectly like other methods for datasets where the number of items was five. This success of EGA was valid even for data sets with a sample size as small as 250. A similar finding was obtained for EGA (TMFG). On the other hand, as the number of items increases, the performance of both EGA (LASSO) and EGA (TMFG) decreased. Even when the sample size was 3000 and the reliability level was 0.9, EGA (TMFG) could not extract the correct number of dimension with high accuracy if there were ten or more items in the data set. On the other hand, for n = 500 and n = 1000 sample size conditions, EGA (LASSO) yielded comparable accuracy rates only if the reliability level was 0.9 while it's performance decreased when the reliability dropped to 0.7 and when data sets contained twenty items.

If the methods are compared in general, AF had perfectly extracted the actual dimensional structure regardless of the conditions altered and use of it by the researchers is strictly recommended in their future studies. Overall, EGA (LASSO) algorithm outperformed EGA (TMFG) algorithm. For this

reason, it is recommended that GLASSO algorithm should be preferred over TMFG algorithm for unidimensional and polytomous data sets. The same superior performance of EGA (GLASSO) was also observed when compared with the traditional KR1 method.

Therefore, it can be said that EGA (LASSO) is an important effective alternative for researchers who prefer the traditional KR1 method, which has been used extensively because of availability on most of the commercial software programs. Considering the richness of output EGA provides (see Golino & Christensen, 2020), EGA can be a better alternative to KR1. In addition, if the sample size was increased to 1000 or 3000, EGA (LASSO) method gives results comparable to the OC and PA methods. On the other hand, EGA should be considered as a serious alternative only when the scale contains fewer items with high internal consistency for smaller sample size conditions (250 or 500). Otherwise, OC and PA provide better results.

According to factorial ANOVA results, it was found that there were no unique or interaction effects observed for EGA (LASSO) method. Similar findings were also observed for OC and PA methods. It can be said that these three methods were the most robust ones across the conditions tested. Although these statistics can not be calculated for AF, it provides perfect results under all conditions. it is also definitely correct to consider this method as robust. On the other hand, "large" effect size was observed for the EGA (TMFG) method for the sample size condition. That is, the sample size affects the performance of EGA (TMFG) method negatively regardless of other conditions. The poor performance of TMFG algorithm is understandable because it performs better when booting algorithms are used simultaneously.

Finally, for the KR1 method. "large" effect sizes were observed for all conditions and their two-way and three-way interactions. Accordingly, it can be said that the KR1 method was the least robust method within the context of the conditions examined in this study. This finding is in line with past literature (Velicer, Eaton & Fava. 2000; Ruscio & Roche, 2012).

This study is one of the few studies comparing EGA's factor retention effectiveness with other traditional methods. Contrary to the findings obtained by Golino and Epskamp (2017), EGA(LASSO) was not to be detected as clearly superior to other traditional methods. This result implies that EGA (LASSO) may not be a suitable alternative when the data is unidimensional and potential researchers should use EGA (LASSO) for scales with fewer items, higher internal consistency and a large sample size for unidimensional tests. On the other hand, EGA (TMFG) should not be an option for researchers in a wide of conditions considered in the current study.

All in all, more research is needed to examine the effectiveness of EGA in different conditions. For example, EGA's effectiveness in datasets with different ability distributions will contribute to the richness of the existing literature. In addition, in this study the effectiveness of the methods was only evaluated in terms of the number of factors. In future studies, it is suggested to evaluate the performance of EGA in terms of estimating real factor loadings.

REFERENCES

Barabási, A.-L., & Pósfai, M. (2016). *Network science*. Cambridge: Cambridge University.

Beard, C., Millner, A. J., Forgeard, M. J., Fried, E. I., Hsu, K. J., Treadway, M. T., Leonard, C. V., Kertz, S. J., & Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological medicine*, *46*(16), 3359–3369. doi: 10.1017/S0033291716002300

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *10*(9), 91-121. doi: 10.1146/annurev-clinpsy-050212-185608

Borsboom, D., & Molenaar, D. (2015). Psychometrics. In J. Wright (Ed.). *International encyclopedia of the social & behavioral sciences* (Second ed.. Vol. 19. pp. 418-422). Amsterdam: Elsevier.

Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review*, *63*(5), 277–293. doi: 10.1037/h0046049

Cattell R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, *1*(2), 245–276. doi: 10.1207/s15327906mbr0102_10

_____

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.

Chalmers, R.P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1-29. doi: 10.18637/jss.v048.i06

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759-771. doi: 10.1093/biomet/asn034

Cliff, N. (1988). The eigenvalue-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*(2), 276-279.

Cohen J, (1988). *Statistical power analysis for the behavior science*. Lawrance Eribaum Association.

de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University.

DeVellis, R. F. (2017). *Scale development: Theory and applications*. Thousand Oaks. CA: SAGE Publications.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, *49*(1), 71-75. doi: 10.1207/s15327752jpa4901_13

DiFranza, J. R., Savageau, J. A., Rigotti, N. A., Fletcher, K., Ockene, J. K., McNeill, A. D., Coleman, M., & Wood, C. (2002). Development of symptoms of tobacco dependence in youths: 30 month follow up data from the DANDY study. *Tobacco control*, *11*(3), 228–235. doi: 10.1136/tc.11.3.228

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of α estimation. *British Journal of Psychology*, *105*(3), 399-412. doi: 10.1111/bjop.12046

Eaton, N. R. (2015). Latent variable and network models of comorbidity: toward an empirically derived nosology. *Social Psychiatry and Psychiatric Epidemiology*, *50*(6), 845-849. doi: 10.1007/s00127-015-1012-7

Edwards, J.R., & Bagozzi, R.P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155-174. doi: 10.1037/1082-989X.5.2.155.

Epskamp, S., & Fried, E. I. (2016). *A primer on estimating regularized psychological networks* arXiv preprint Stat-Ap/1607.01367. Available at: http://arxiv.org/abs/1607.013677

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195-212. doi: 10.3758/s13428-017-0862-1

Epskamp, S., Maris, G., Waldorp, L.J., & Borsboom, D. (2015). Network Psychometrics. In Irwing. P., Hughes, D. and Booth, T. (Eds.). *Handbook of Psychometrics*. New York: Wiley.

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of abnormal psychology*, *126*(8), 1044–1056. doi: 10.1037/abn0000311

Foygel, R. and Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems*, 23, 2020-2028.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, *9*(3), 432-441. doi: 10.1093/biostatistics/kxm045

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*(1). 93–111. doi: 10.1037/met0000064

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035. doi: 10.1371/journal.pone.0174035

Golino, H., Shi. D., Christensen, A. P., Garrido, L. E.. Nieto, M. D., Sadana, R., Thiyagarajan. J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*. Advance online publication. doi: 10.1037/met0000255

Golino, H.. & Christensen, A. P. (2020). *EGAnet: Exploratory Graph Analysis -- A framework for estimating the number of dimensions in multivariate data using network psychometrics*. R package version 0.9.4.

Gorsuch R.L. (1988) Exploratory Factor Analysis. In: Nesselroade J.R., Cattell R.B. (eds) *Handbook of Multivariate Experimental Psychology. Perspectives on Individual Differences*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-0893-5_6

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149-161. doi: 10.1007/BF02289162

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185. doi: 10.1007/BF02289447

Kline, P. (2014). *An easy guide to factor analysis*. Routledge.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202–220. doi: 10.1177/1094428105284919

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

10

Lauritzen, S. L. (1996). *Graphical Models. Oxford Statistical Science Series*. volume 17. New York. NY: Oxford University Press.

Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated Maximally Filtered Graph. *Journal of Complex Networks*, 5, 161–178. doi: 10.1093/comnet/cnw015

Nunnally. J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University.

Pons, P.. & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms Applications, 10*(2), 191-218. doi: 10.1007/11569596_31

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. Austria. URL https://www.R-project.org/.

Radloff, L.S. (1977). The CES–D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385-401. doi: 10.1177/014662167700100306

Raiche, G. (2010). *nFactors: An R package for parallel analysis and non graphical solutions to the Cattell's scree test*. R package version 2.3.3.

Raiche, G., Riopel. M. and Blais, J.-G. (2006). *Non graphical solutions for the Cattell's scree test*. Paper presented at the International Annual Meeting of the Psychometric Society, Montreal.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403-414. doi: 10.1207/s15327906mbr1404_2

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282-292. doi: 10.1037/a0025697

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. New Ideas in Psychology, *31*(1), 43-53. doi: 10.1016/j.newideapsych.2011.02.007.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327. doi: doi.org/10.1007/BF02293557

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.). *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (p. 41-71). Kluwer Academic/Plenum Publishers. doi: 10.1007/978-1-4615-4397-8_3

Tek Boyutlu ve Çok Yanıt Kategorisine Sahip Veriler İçin Açıklayıcı Grafik Analizinin Performansının İncelenmesi

Giriş

Gözlenen değişkenlerin örtük yapılarla nasıl ilişkilendirilmesi gerektiği sorusu psikometrinin merkezinde yer almaktadır (Borsboom ve Molenaar, 2015). Şimdiye kadar, bu ilişkiyi belirtmek için çeşitli modeller geliştirilmiştir. Bununla birlikte, psikometrik çalışmalarda kullanılan matematiksel modellerin niceliksel artışına ve büyük esnekliğine rağmen, örtük özellikler ve davranışlar arasındaki ilişkileri tanımlamak için sunulan modeller, dayandıkları paradigma açısından şaşırtıcı bir şekilde sınırlıdır.

Bu geleneksel yaklaşımlara ise yakın zamanda ağ modellemesi olarak adlandırılan alternatif model önerilmiştir. Bu yaklaşımda, gözlenen değişkenler arasındaki ilişkilerin, değişkenlerin birbirleriyle doğrudan ve potansiyel olarak nedensel etkileşime sebep olan bir sistem aracılığıyla kaynaklandığı varsayılmaktadır (Eaton, 2015). Ağ modellerinin kullanımı, birçok farklı disiplindeki karmaşık sistemlerin anlaşılması için büyük ölçüde fayda sağlamıştır (Barabási ve Pósfai, 2016). Sosyal bilimlerde, ağ analizi uygulaması öncelikle sosyal ağ yapılarını araştırmak için benimsenmiştir (örn. Cartwright ve Harary, 1956). Bununla birlikte, sonraki yıllarda, psikolojik davranışların ağ modellerini keşifsel bir şekilde analiz edilmesi geleneksel gizli değişken modellemelerine alternatif olarak kullanılmaya başlamıştır (Borsboom ve Cramer, 2013; Schmittmann vd., 2013). Ağ modelleme uygulamasındaki bu değişimden sonra, ağ yaklaşımının popülaritesi artmış ve psikoloji alanında

_____

yoğun bir şekilde kullanılmaya başlanmış ve psikolojik verilerde ağ yapılarını tahmin etmeyi amaçlayan yeni bir psikoloji alanının ortaya çıkmasına neden olmuştur. Bu yeni alan, ağ psikometrisi olarak adlandırılır (Epskamp, Maris, Waldorp ve Borsboom, 2015).

Diğer ağ modellerinde olduğu gibi, psikometrik ağ modeli de bir dizi düğümden (veya köşelerden), düğümler arasında bir dizi bağlantı veya ağdan (kenarlar olarak da bilinir) ve düğümlerin ve kenarların yapısıyla ilgili bilgilerden oluşur (De Nooy, Mrvar ve Batagelj, 2011). Düğümler, psikolojik gösterge değişkenlerini (örn. gizil değişkenlerin semptomları, davranışları veya yüzleri) temsil eder. Geleneksel olarak, düğümler ağ yapısında dairelerle temsil edilirler. Öte yandan, kenarlar, düğümler arasındaki ilişkileri temsil eder ve bir ağ modelinde daireleri birbirine bağlayan çizgilerle temsil edilir.

Yakın geçmişte yayımlanan bir çalışma (Golino ve Epskamp, 2017), ağ modelleme yoluyla psikolojik yapıların boyutluluğunu araştırmanın yenilikçi bir yolunu sunmuştur. Bu yeni tekniğe Açıklayıcı Grafik Analizi (AGA) adı verilir. Adından da anlaşılacağı gibi, bu model bir yapıyı incelerken önsel varsayımlara dayanmamaktadır. Bunun yerine, psikolojik verileri keşifsel bir anlayışla ele alır. AGA'nın dikkate değer bir özelliği, değişkenler arasındaki ilişkilerin görselleştirilmesi ve değişkenlerin boyutlara atanmasını belirleyici bir şekilde sağlamasıdır. Bu nedenle psikolojik özelliklerin kuramsal yapısını test etmek veya yeniden değerlendirmek için ideal bir yöntemdir.

Kısmi korelasyonlara dayalı tahminlerle gerçekleştirilen bu yöntemde, yaygın olarak en az mutlak daralma ve seçim operatörünün (least absolute shrinkage and selection operator-LASSO) işlemi yaygın olarak uygulanmaktadır (Friedman, Hastie ve Tibshirani, 2008). LASSO, sahte (spurious) bağlantıları kontrol etmek için kullanılmaktadır. LASSO, walktrap gibi topluluk algılama algoritmalarıyla birleştirildiğinde yüksek hassasiyetli tahminler sağlayabilir (Pons and Latapy, 2005).

Optimum bir model elde etmek için Chen ve Chen (2008) tarafından önerilen genişletilmiş Bayesian bilgi kriteri (extended Bayesian information criterion-EBIC) dikkate alınarak belirlenen ayarlama parametresi kullanılır. Bu parametre, araştırmacının ağların seyrekliğini kontrol etmesini sağlar (Foygel ve Drton, 2010). LASSO, nihai ağ yapısını belirlediği için ağ modellemenin önemli bir parçasıdır. Aynı zamanda daha tutucu ve yorumlanabilir modellerin elde edilmesini sağlar. AGA modellerinde LASSO'nun grafik bir uzantısı kullanılmış ve GLASSO olarak adlandırılmıştır. Ek olarak, Üçgenleştirilmiş Maksimum Filtrelenmiş Grafik (TMFG: Triangulated Maximally Filtered Graph), GLASSO'ya alternatif olarak önerilen bir diğer tekniktir. Bu yaklaşım, bir puan fonksiyonunun maksimize etmesini sağlayan bir üçgenleme oluşturur. Bu şekilde veriler anlamlı bir yapı içerisinde organize olur ve modelleme mümkün olur. Ayrıntılı açıklamalar ve formülasyonlar için Massara, Di Matteo ve Aste (2016) 'ye bakılması önerilmektedir.

Bir AGA modelinde, geleneksel olarak, ağ üzerindeki yeşil (ya da mavi) çizgiler pozitif kısmi korelasyonları temsil ederken kırmızı çizgiler, negatif kısmi korelasyonlara karşılık gelir. Ek olarak, çizgilerin kalınlığı korelasyon miktarı hakkında bilgi verir: daha kalın çizgiler kısmi korelasyon değerlerinin 1'e yaklaştığını gösterir, Kısmi korelasyon değerleri tam olarak 0 ise, iki düğüm arasında hiçbir çizgi çizilmez. Yani, ağdaki diğer değişkenlerin etkisi kontrol edildiğinde iki değişken koşullu olarak bağımsızdır (Pearl, 2000).

AGA'nın önerildiği makalede Golino ve Epskamp (2017), AGA'nın performansını beş farklı geleneksel faktör çıkarma tekniğiyle karşılaştırmıştır. Bu çalışmada iki yanıt kategorili maddelerden oluşan iki ve dört boyutlu türetilmiş veri setleri kullanılmıştır. Çalışmanın bulguları, kontrol edilen koşullar ne olursa olsun, özellikle veri kümeleri dört boyutlu yapı olarak simüle edildiğinde AGA'nın en iyi performans gösteren yöntem olduğunu göstermiştir. Özellikle boyut sayısı dört olduğunda, AGA'nın diğer geleneksel yöntemlere üstünlüğünü doğrulanmıştır. Özellikle AGA'nın her boyuttaki madde sayısı beş olduğunda tatmin edici sonuçlar veren tek yöntem olduğu belirtilmiştir. Bu çalışmada çok boyutlu veriler kullanılmış olmasına ragmen daha sonrasında AGA algoritması tek boyutlu veri setlerinin incelenmesine izin verecek şekilde revize edilmiştir. Nitekim, aynı çalışmada AGA'nın tek boyutlu veri setleri için faktör sayısına karar vermedeki performansının incelenmesi önerilmiştir. Bu öneri dikkate alınarak gerçekleştirilen bu çalışmanın amacı veri seti tek boyutlu olduğunda ve maddeler çok yanıt kategorisine sahip olduğunda AGA'nın faktör sayısına karar vermedeki performansının geleneksel faktör çıkarma yöntemleriyle karşılaştırılması olarak belirlenmiştir.

_____
ISSN: 1309 – 6575   _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
Journal of Measurement and Evaluation in Education and Psychology

12

Yöntem

Bu çalışmada üç farklı koşul kontrol edilmiştir: örneklem büyüklüğü (250, 500, 1000 ve 3000), madde sayısı (5, 10, 20) ve iç tutarlılık seviyesi ($\alpha = 0.7$ ve $\alpha = 0.9$). Buna koşullara bağlı olarak 4x3x2 tasarımı ile 24 farklı koşul oluşturulmuştur. Ayrıca, bu çalışmanın temel amacı doğrultusunda, tüm veri setleri tek boyutlu yapıya sahip olacak şekilde türetilmiştir ve maddeler 1-5 aralığında puanlanmış şekilde veri setleri oluşturulmuştur. Daha kararlı sonuçlar elde etmek için her koşul için veri üretme işlemi 100 kez tekrarlanmıştır. Bu sayede, 2400 veri kümesi türetilmiştir. Bu çalışmada bulgular kısmında sunulan sonuçlar, tekrarlar sonucunda elde edilen değerlerin aritmetik ortalamasını yansıtmaktadır. Veri üretme işlemi, R ortamında (R çekirdek ekibi, 2019) "mirt" paketi (Chalmers, 2012) ile gerçekleştirilmiştir.

AGA yönteminin performansını karşılaştırmak amacıyla beş farklı faktör sayısına karar verme yöntemi kullanılmıştır: Hızlanma Faktörü (AF: Acceleration Factor), Optimal Koordinatlar (OC: Optimal Coordinates), Paralel Analiz (PA) ve Kaiser'in özdeğer 1'den büyük kuralı (KR1). Bu dört yönteme ilişkin ayrıntılı teknik bilgiler Raiche, Riopel & Blais (2006) ve Raiche (2010) 'de yer almaktadır. AGA analizleri, R istatistik programında bulunan "EGAnet" paketi (Golino & Christensen, 2020) kullanılarak gerçekleştirilirken, OC, AF, PA ve KR1 faktör çıkarma yöntemleri için "nFactors" paketi (Raiche, 2010) kullanılmıştır. AGA tekniği GLASSO ve TMFG algoritmaları için ayrı ayrı gerçekleştirilmiştir ve çalışmanın geri kalanında sırasıyla AGA(GLASSO) ve AGA(TMFG) anılmıştır. GLASSO algoritması kullanılırken, ayarlama parametresi 0.5 olarak belirlenmiştir. Koşullara göre genel betimleyici istatistiklerin yanısıra faktöriyel varyans analizi (ANOVA) gerçekleştirilerek etkisi incelenen koşulların faktör sayısına karar verme yöntemleri üzerindeki tekil etkileri ile etkileşimlerinden kaynaklı etkilerin incelenmesi amaçlanmıştır.

Garrido, Abad ve Posada (2016) tarafından önerildiği gibi, doğru boyut sayısının ne kadar kesinlikte çıkarıldığına dair değerlendirme, çıkarma doğruluk indeksi ve yanlılık indekslerine dayanılarak yapılmıştır. Faktör çıkarma doğruluk indeksi, doğru sayıda faktörün 1, hatalı sayıda faktörün 0 olarak çıkarıldığı analiz sonuçlarının kodlanmasıyla elde edilmiştir. Örneğin 100 veri seti incelendiğinde, 50 veri seti için gerçek faktör sayısı çıkarılmışsa bu veri setlerinin her biri 1, geri kalanı ise 0 olarak kodlanmıştır. Sonuç olarak 100 veri seti için yöntemin nihai kesinlik 0.5 olarak hesaplanmıştır. Öte yandan, yanlılık indeksi, kestirilen boyut sayısının gerçek boyut sayısından çıkarılmasıyla hesaplanır. Örneğin, tek boyutlu bir veri seti için, kestirilen boyut sayısı 1 ise, sapma endeksi 0 olarak hesaplanırken, kestirilen boyut sayısı 2 ise yanlılık değeri 1 olur. Başka bir anlatımla, sıfır yanlılık değeri, boyut sayısının doğru kestirildiğini gösterirken 0'dan uzak yanlılık değerleri, ilgili yöntemin zayıf performansını göstermektedir.

Sonuç ve Tartışma

Elde edilen bulgulara göre AGA (LASSO) madde sayısının 5 olduğu veri setleri için diğer yöntemler gibi tek boyutlu yapıyı mükemmel bir şekilde kestirdiği görülmüştür. AGA'nın bu başarısı, örneklem büyüklüğü 250 olan veri setleri için bile geçerlidir. Benzer bulgular AGA (TMFG) için de elde edilmiştir. Diğer taraftan, madde sayısı arttıkça hem AGA (LASSO) hem de AGA (TMFG)'nın performansının düştüğü görülmüştür. Örneklem büyüklüğü 3000 ve güvenilirlik düzeyi 0.90 olsa bile AGA (TMFG) veri setinde 10 veya daha fazla madde olduğunda doğru boyut sayısını yüksek doğrulukla çıkartamadığı belirlenmiştir. Ayrıca, n = 500 ve n = 1000 örneklem büyüklüğü koşulları için AGA (LASSO) yalnızca güvenilirlik seviyesi 0.9 olduğunda diğer yöntemlerle karşılaştırılabilir kesinlik oranları sağlamıştır. Ancak, güvenilirlik 0.7'ye düştüğünde ve veri kümeleri 20 madde içerdiğinde performansı düşmüştür.

Yöntemler genel olarak karşılaştırıldığında ise. AF'nin control edilen koşullardan bağımsız olarak gerçek boyutsal yapıyı mükemmel bir şekilde çıkarttığı ve gelecekteki çalışmalarında araştırmacılar tarafından tercih edilebileceği görülmüştür. Genel olarak. AGA (LASSO) algoritması, AGA (TMFG) algoritmasından daha iyi performans göstermiştir. Bu nedenle, tek boyutlu ve çok yanıt kategorisine

_____

sahip veri setleri için GLASSO algoritmasının TMFG algoritmasına tercih edilmesi gerektiği görülmüştür. AGA (GLASSO)'nın üstün performansı, geleneksel KR1 yöntemiyle karşılaştırıldığında da gözlenmiştir. Bu nedenle, AGA (LASSO)'nın geleneksel KR1 yöntemini tercih eden araştırmacılar için önemli ve etkili bir alternatif olduğu söylenebilir. AGA'nın sağladığı bilgilerin zenginliği göz önüne alındığında, araştırmacılar tarafından tercih edilmesi özellikle önerilmektedir. Ek olarak, örneklem büyüklüğü 1000 veya 3000'e yükseltildiğinde AGA (LASSO) yöntemi, OC ve PA yöntemleriyle de karşılaştırılabilir sonuçlar vermiştir. Örneklem büyüklüğü daha küçük ise (250 veya 500), AGA yalnızca yüksek iç tutarlılığa sahip ve daha az madde içeren ölçüm araçları için ciddi bir alternatif olarak düşünülmelidir.

Faktöriyel ANOVA sonuçlarına göre AGA (LASSO) yöntemi için tekil veya etkileşim etkisinin gözlemlenmediği bulunmuştur. OC ve PA yöntemleri için de benzer bulgular gözlemlenmiştir. Bu üç yöntemin test edilen koşullar arasında en dayanıklı yöntemler olduğu söylenebilir. Ayrıca bu istatistikler AF için hesaplanamamıştır çünkü bu yöntem her koşulda mükemmel sonuçlar ortaya koymaktadır. Başka bir anlatımla, bu yöntemi sağlam yöntem olarak değerlendirmek mümkündür. Öte yandan, örneklem büyüklüğü koşulunun AGA (TMFG) yönteminde "büyük" etkiye sahip olduğu gözlemlenmiştir. Yani örneklem büyüklüğü AGA (TMFG) yönteminin faktör sayısına karar verme performansı üzerinde etkiye sahiptir. Son olarak, geleneksel olarak en yaygın kullanılan KR1 yöntemi için tüm koşullar ve bunların ikili ve üçlü etkileşimleri için "büyük" etkiler gözlemlenmiştir. Buna göre bu çalışmada incelenen koşullar bağlamında KR1 yönteminin en az sağlam yöntem olduğu söylenebilir. Bu bulgu, ilgili alan yazın ile uyumludur (Velicer. Eaton. Fava. 2000; Ruscio & Roche. 2012).

Bu çalışma, AGA' nın faktör çıkarma etkinliğini diğer geleneksel yöntemlerle karşılaştıran birkaç çalışmadan biridir (Golino & Epskamp. 2017; Golino ve ark. 2020). Bu nedenle, AGA' nın farklı koşullarda etkinliğini incelemek için daha fazla araştırmaya ihtiyaç vardır. Örneğin, farklı yetenek dağılımlarına sahip veri setlerinde AGA' nın etkinliği, mevcut alan yazının zenginliğine katkıda bulunacaktır. Ayrıca, bu çalışmada yöntemlerin etkinliği yalnızca faktör sayısını kesin ve yansız çıkartabilme açısından değerlendirilmiştir. Gelecek çalışmalarda, AGA'nın performansının gerçek faktör yüklerini tahmin etme açısından değerlendirilmesi önerilmektedir.

_____

ISSN: 1309 – 6575   _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
Journal of Measurement and Evaluation in Education and Psychology   14