Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN Online : 2312-9883
QJAE, Volume 27, Issue 1 (2025) ISSN Print : 1816-9171
DOI: https://doi.org/10.33916/qjae.2025.012734

Studying The Factors Affecting divorce using Regularization technique of two stages high- order in factor analysis

Mohammed H. AL-Sharoot Rammah Oday Hassan
mohammed.alsharoot@qu.edu.iq rimah.dunbl@uobabylon.edu.iq
University of Al-Qadisiyah

Article history:
Received: 22/12/2024
Accepted: 24/12/2024
Available online: 25 /3 /2025

Corresponding Author : Rammah Oday Hassan

Abstract : This study investigates the application of regularization methods within a two-stage high-factor analysis framework. This proposed method aims to handle high-dimensional and complex data. First, we reduce the high factors (dimensionality) of the model by using factor analysis to find important factors via regularization techniques to improving of the interpretability and stability of the model. This stage involves the development of using the selected factors, via Lasso are applied to mitigate overfitting and improve generalization, the proposed method including regularization not only speeds up the factor selection process but also produces predictions that are more accurate. But the second stage is focus on regularization the variables within the factor, to improve the interpretability and stability of the model . Also ,we can minimize overfitting and enhance the model's performance in high-dimensional data scenarios by regularizing these variables efficiently. The superiority of our proposed method compared to existing approaches in this field can be demonstrated using simulation techniques and real-world data. Where, we notice that the proposed method has proven its superiority over previous methods in the same field, and this is evident from the presented Results.

Keywords: high factor analysis, regularization methods, regularizing important factor, selection variables

INTRODUCTION: A statistical method called factor analysis divides the observable variables into smaller unobserved variables called factors (Kim, J. O., & Mueller, C. W. (1978)). The number of extracted factors will equal the number of variables, with each factor representing a linear combination of these variables. This means that each factor represents a set of observed variables that contribute to the variation in the data Costello, A. B., & Osborne, J. W. (2005). Numerous criteria exist for identifying the most significant factors; the Kaiser criterion(Kaiser, H. F. (1960)), which emphasizes choosing factors that correspond to eigenvalues larger than one, is among the most widely used approaches Field, A. (2013). However, in the current study , we will focus on a regularization method that balances bias and variance on one hand, it know Lasso technique, while reducing time and improving the interpretability of the model by automatically selecting important factors and setting insignificant factors to zero exactly(Tibshirani, R. (1996)). The majority of the time, when reducing high-dimensional by direct factor analysis, the remaining factors have low explanatory ability , as compered by the explained variance to total variance ratio. To overcome this issue, high factor analysis can be used, which represents the optimal level of analysis that achieves the highest explained variance (Cieciuch, J, *et al* ,,2014). In this paper, we mixed between lasso with high factor analysis, Where the model becomes simpler to understand when fewer factors are included, which frees up the researchers' attention to concentrate on the most important factors that influence the result. Also, this mixing can lead to improved model performance by reducing overfitting and enhancing generalization to new data. Additionally, by lessening overfitting and boosting generalization to a new data, this mixing can increase model performance. This combination can save time and computational resources by automating the identification of significant factors and important variables selection within the factors, hence streamlining the analytic process. This paper organized

High-order factor analysis

One crucial technique in multivariate analysis is factor analysis. It is mostly focuses with examining the interrelationships (correlations) among a set of variables(Henson, R. K. *et al* ,,2006). This a set variables $(x_1, x_2, \dots x_n)$ are related to one another either directly or indirectly, and factor analysis reduces them to a smaller set of variables known as common factors$(Z_1, Z_2, \dots Z_k)$. Following that, these factors are modeled with specific linear models that preserve a substantial amount of information from the initial variables(Okon, Jan(1974)). We can used the matrix system for writing mathematical model to factor analysis as following :

$$Z = \gamma X + \varepsilon \qquad (1)$$

where X is original variables with $(n * 1)$, γ is loading factor (correlation coefficients between observed variables and latent factors in dimensions $(k * n)$,.)
$Z$is Common factors values with $(k * 1)$, ε is residual (error) values in dimensions$(k * 1)$.
Additionally, we shall impose the subsequent assumptions on Z
-Common factors and error are independent $cor(Z, \varepsilon) = 0$, This indicated that the errors are unrelated to the common factors.
- $E(Z) = 0$, may have significant effects on statistical modeling by implying that common factors do not always favour positive or negative values, which results in more thorough analysis. This indicated that the errors are unrelated to the common factors.
- $Cov(Z) = I$, where is the identity matrix, I is the cov covariance matrix, and is used to ensure that the common factors are uncorrelated.
Let $Cov(X) = \Sigma$,
$\Sigma = \mathrm{Cov(X)} = \mathrm{Cov}(\gamma Z + \varepsilon)$
$\mathrm{Cov}(\gamma Z + \varepsilon) = Cov(\gamma Z) + \mathrm{Cov}(\varepsilon)$,From the above conditions $E(\gamma Z) = \gamma E(Z) = 0$
$\Sigma = \gamma Cov(Z)\gamma^T + \mathrm{Cov}(\varepsilon)$,let $\mathrm{Cov}(\varepsilon) = \varpi$
$\Sigma = \gamma\gamma^T + \varpi$
From the model presented in equation (1), we find that the factor loadings (γ) summarize the relationships between the variables and the original factors.

$$Cor\left(X_i, Z_j\right) = \gamma_{ij} \qquad (2)$$

In factor analysis, γ_{ij} is matrix of factor loading is a crucial strategy that aids in the process of dimensionality reduction and enables researchers to comprehend patterns of correlations between variables and underlying factors.Factors after rotation may depend on other factors, which is referred to as second-order factor analysis. Since factor analysis reduces a large set of variables to a smaller number of factors, these factors can be considered as variables themselves, allowing for a second round of factor analysis (Cudeck, R. (2000). Often, the explained variance from first-order factors (direct analysis) is low and does not represent the explanatory power of the phenomenon under study. To overcome this issue, we used the higher-order factor analysis. A statistical technique known as higher-order factor analysis involves factor analysis, oblique rotation, and factor analysis of rotated factors as successive processes. Its advantage is that it makes the hierarchical structure of the phenomena under study visible to the researcher. One can either post-multiply the primary factor pattern matrix by the higher-order factor pattern matrices in order to comprehend the results (Gorsuch, 1983). To achieve the optimal level in factor analysis, we rely on the degree that achieves the highest explained variance, which allows us to focus and reduce factors to those with high explanatory power. To implement second-order factor analysis , the model will follow the following form:

$$Z^1 = \gamma^1 X + \varepsilon \qquad (3)$$

where X is original variables with $(n * 1)$, γ^1 is loading factor (correlation coefficients between observed variables and second –order latent factors in dimensions $(r * n)$,.) $Z^1$is second –order Common factors of s with $(r * 1)$, ε is residual (error) values in dimensions$(r * 1)$. The second-order factor loadings represent the relationship between the original variables and the second-order factors after projecting the original variables onto the second-order factors, as shown in the following equation.

$$Cor\left(X_i, Z_j^1\right) = \gamma_{ij}^1 \qquad (4)$$

To implement third-order factor analysis , the model will follow the following form:

$$Z^2 = \gamma^2 X + \varepsilon \qquad (5)$$

where X is original variables with $(n * 1)$, γ^2 is loading factor (correlation coefficients between observed variables and third –order latent factors in dimensions $(l * n)$,.) $Z^2$is third –order Common factors of s with $(l * 1)$, ε is residual (error) values in dimensions$(l * 1)$. The second-order factor loadings represent the relationship between the original variables and the second-order factors after projecting the original variables onto the third-order factors, as shown in the following equation (Taha .H. A, and Fadhel H. H,2010).

$$Cor\left(X_i, Z_j^2\right) = \gamma_{ij}^2 \qquad (6)$$

To implement high-order factor analysis , the model will follow the following form:

$$Z^h = \gamma^h X + \varepsilon \qquad (7)$$

where X is original variables with $(n * 1)$, γ^h is loading factor (correlation coefficients between observed variables and high –order latent factors in dimensions $(b * n)$,.) $Z^h$is high –order Common factors of s with $(b * 1)$, ε is residual (error) values in dimensions$(b * 1)$. The high-order factor loadings represent the relationship between the original variables and the second-order factors after projecting the original variables onto the high-order factors, as shown in the following equation.

$$Cor\left(X_i, Z_j^h\right) = \gamma_{ij}^h \qquad (9)$$

In this paper, we introduced a good method via mixing one of regularization technique (lasso) and high-order factor analysis.

Regularization technique of high- order factor analysis

High-order factor analysis and regularization **t**echnique (lasso) together can improve the modeling process in a number of ways. An outline of the associated methods and benefits is provided below. This proposed new approach will provide us with a set of advantages. First, it will uniquely contribute to selecting important factors by excluding insignificant factors through setting their eigenvalues to zero.

$$Z^h = \gamma^h X + \lambda \sum_{j=1}^{k} \left|\gamma_{ij}^h\right| \qquad (10)$$

$\lambda \geq 0$, is a regularization loading factor that controls the strength of the penalty applied. $\sum_{j=1}^{k}\left|\gamma_{ij}^h\right|$ As is common in regularization procedures to promote sparsity in the model, this term sums the absolute values of the loading factor. Second , it will uniquely contribute to selecting important variable within the important factor as the second stage by excluding insignificant factors through setting some elements of eigenvector to zero exactly. The following mathematical model illustrates the selection of variables within a single factor.

$$Lof\ F = Z^h \sqrt{\gamma^h} + \lambda \sum_{j=1}^{k} |\gamma^h| \qquad (11)$$

LoF is stand for the factor loadings, or factor loadings matrix((Toczydlowska, D et al 2017).
Z^h is stand high-order factor analysis, or eigenvector matrix.
$\sqrt{\gamma^h}$ is represents the eigenvalue's square root for the high-order factors more detail ((Jolliffe, I et al 2003)) .
λ is represented the regularization parameter for high-order factor .
The term of $\lambda \sum_{j=1}^{k} |\gamma^h|$In regularization techniques, this term sums absolute values of the loading factors , which promotes sparsity in the model. (Tibshirani, R. (1996)).
In our proposed method, the code from the (psych) code has been modified by integrating Lasso technique for selecting important factors, as well as identifying significant variables within those factors by employing the (glmnet) code.

Real data set

issues that affects individuals, families, and communities as a whole. In recent years, divorce rates have seen a noticeable increase in many cultures and countries, raising questions about the causes and factors associated with this phenomenon. Risks of the Phenomenon Divorce not only affects the spouses but also extends its impact to children, parents, and friends, creating changes in family dynamics and affecting social relationships. Understanding the causes and motivations behind divorce can help develop supportive strategies for couples and families, contributing to the improvement of marital relationships and enhancing family stability. There are Influencing many variables intersect in the phenomenon of divorce, including

1- x_1: $Changes\ in\ values$
2- x_2: $Increased\ independence\ of\ women.$
3- x_3: Financial pressures.
4- x_4:Living conditions.
5- x_5: Personal conflicts..
6- x_6: $Traditions\ and\ beliefs$
7- x_7: $Influence\ of\ friends\ and\ family.$
8- x_8: Age at marriage.
9- x_9: Presence of children.
10-x_{10}: Impact of social media.
11-x_{11}: Second or third marriage.

the questionnaire was presented to 120 social researchers in the courts of the provinces of Hilla and Qadisiyyah. After collecting the questionnaires from their original sources, it was found that there were 8 invalid questionnaires that were excluded. The valid questionnaires were processed numerically and entered into the program for data analysis.

Results

The above data will be analyzed using the proposed method (lasso high-order factor analysis)and one of the classical approaches (Kaiser criterion) to select important factors and explained variance in both methods, along with a comparison of the results.

Table -1- show the Eigen value and explained variance of our proposed method(Lasso high-order factor analysis) and classical method (factor analysis)for first-order factor analysis

Factors	Fac1	Fac2	Fac3	Fac4	Fac5	Fac6	Fac7	Fac8	Fac9	Fac10	Fac11
Eigen value	1.551	1.463	1.320	1.288	1.159	1.102	1.094	0.753	0.545	0.491	0.234
Explained variance%	13.221	12.783	11.106	10.863	9.657	9.114	8.755	7.526	6.845	5.496	4.634
Proposed method											
Lasso Eigen value	1.967	0.000	0.845	0.782	0.000	0.645	0.000	0.883	0.000	0.000	0.000
lassoExplained variance%	19.341	0.000	17.453	16.673	0.000	16.562	0.000	12.652	0.000	0.000	0.000

From the results presented in the table above, we observe that the number of factors is equal to the number of variables, which is 11. Therefore, the first-order factor analysis was able to identify 7 significant factors out of 11 and disregard 4 insignificant factors, relying on the criterion of eigenvalues greater than one. We note that these 6 significant factors explained 75.499% of the total variance. However, in our proposed method, we find that the number of non-zero factors is 5. This means that the first-order factor analysis successfully identified 5 significant factors out of 11 and excluded 6 factors that were deemed insignificant, setting their coefficients to zero exactly and excluding them from the analysis. Our proposed method automatically selected the important factors by zeroing out the coefficients of the insignificant factors. We also note that the significant factors were able to explain 85.312% of the total variance.

. Therefore, the important variables within the important factors for classical methods (factor analysis) , we can show in the following figure

Figure -1- Show the factor loading of original variables with important first –order factor by classical method

As mentioned earlier, the significant first-order factors are 7 factors, each containing original variables with different factor loadings. The blue color corresponds to the significant variables within the factors that have factor loadings greater than 0.5. In contrast, the yellow color represents the insignificant variables within the factors that have factor loadings less than 0.5. the below figure show non-zero and zero loading factor for each factor

Figure -2- Show the factor loading of original variables with important first –order factor by proposed method (Lasso high-order factor analysis)

From the above figure, we observe that the non-zero factor loadings represented in blue correspond to the significant variables within each factor. Conversely, the remaining variables have zero factor loadings, indicating that these variables do not have an effect on their studied factors and can be excluded.

Via the observing both methods, we find that there is a percentage of unexplained variance .The classical method was unable to explain approximately 24.501% of the total variance, while our proposed method failed to explain 14.688% of the total variance. Therefore, first-order factor analysis cannot be considered the optimal degree. we see that there is unexplained variance in both methods at different rates. Therefore, it cannot be concluded that first-order factor analysis represents the optimal degree of analysis until the explained variance of the second-order factors is calculated and compared with the explained variance rates of the first and second-order factors. The process of second-order factor analysis for both methods relies on the factor loadings of the original variables with the first-order factors, as shown in the figures(1)(2). These are treated as original variables for the subsequent analysis steps, as illustrated in the table below.

Table -2- Show the Eigen value and explained variance of our proposed method(Lasso high-order factor analysis) and classical method (factor analysis)for second-order factor analysis

Second –order Factors	Fac_1^1	Fac_2^1	Fac_3^1	Fac_4^1	Fac_5^1	Fac_6^1	Fac_7^1
Eigen value	1.542	1.393	1.165	1.029	0.716	0.642	0.513
Explained variance%	19.453	18.857	17.762	15.452	12.542	9.673	6.172
Proposed method							
Lasso Eigen value	2.736	0.965	1.0672	0.000	0.000	------	------
Lasso Explained variance%	32.784	30.563	27.654	0.000	0.000	-----	------

From the results of the table above, which pertain to the second-order factor analysis of the classical method, we find that the number of factors with eigenvalues greater than one is 4. Thus, the number of significant factors is 4 out of 7 factors, and these 4 factors were able to explain 71.524% of the total variance. When comparing this percentage with the explained variance in the first-order factors, we see that the explained variance in the second-order factors is lower than that in the first-order factors. Therefore, the analysis can be concluded at the first-order factors, which are considered a good method for analyzing the phenomenon under study. But , we find that the number of non-zero factors in the second-order analysis of the proposed method(lasso high-order factor analysis) is 3 out of a total of 5 factors. These three factors explained variance is 91.001% of the total variance. When comparing this percentage with the explained variance in the first-order factors, we see that the explained variance in the second-order factors is greater than that in the first-order factors. Thus, second-order factor analysis is

superior to first-order factor analysis. We will continue with the analysis steps to reach an appropriate degree of factor analysis to rely on for analyzing the remaining degrees.
In our proposed method (Lasso high-order factor analysis) with second -order factors are 3 factors have non-zero eigenvalue ,each these factors have zero and non-zero factor loading for the first -order factors with second -order factors as show in the following figure

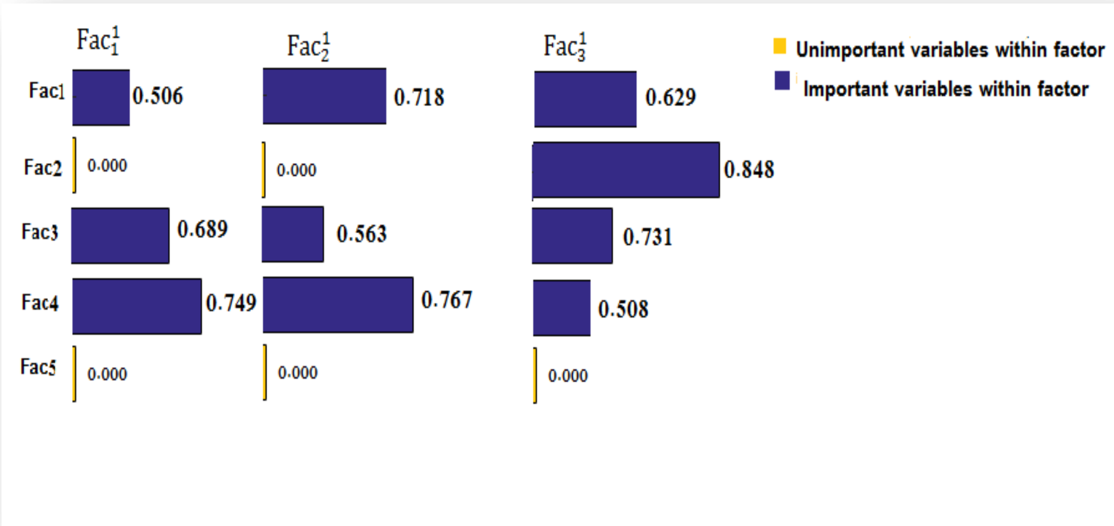

Figure -3- Show the factor loading of first-order factors with important second –order factor by proposed method (Lasso high-order factor analysis)
In our proposed method(lasso), we can continue with third-order factor analysis by relying on the factor loadings of the first-order factors with the second-order factors. The analysis steps can be repeated for the second-order factors to obtain the third-order factors, as illustrated in the table below.

Table -3- Show the Eigen value and explained variance of our proposed method(Lasso high-order factor analysis) third second-order factor analysis

Third –order Factors	Fac_1^2	Fac_1^2	Fac_1^3
Lasso Eigen value	1.673	1.0468	0.000
Lasso Explained variance%	44.745	38.935	0.000

From the results shown in the table above, we find that the non-zero eigenvalues are 2 factors. This means that the number of significant factors is 2 out of 3 factors, and these two important factors were able to explain 83.680% of the total variance. When comparing this variance ratio with the explained variance of the second-order factors, which is greater than this ratio, we conclude that we will rely on the second-order factors based on this results. After determining the second-order factors, which represent the optimal level in the analysis of the phenomenon under study, we notice a separation between the original variables and the second-order factors. The factor loadings of the original variables can be recalculated with the second-order factors by projecting the original variables onto the second-order factors, as illustrated in the following figure:

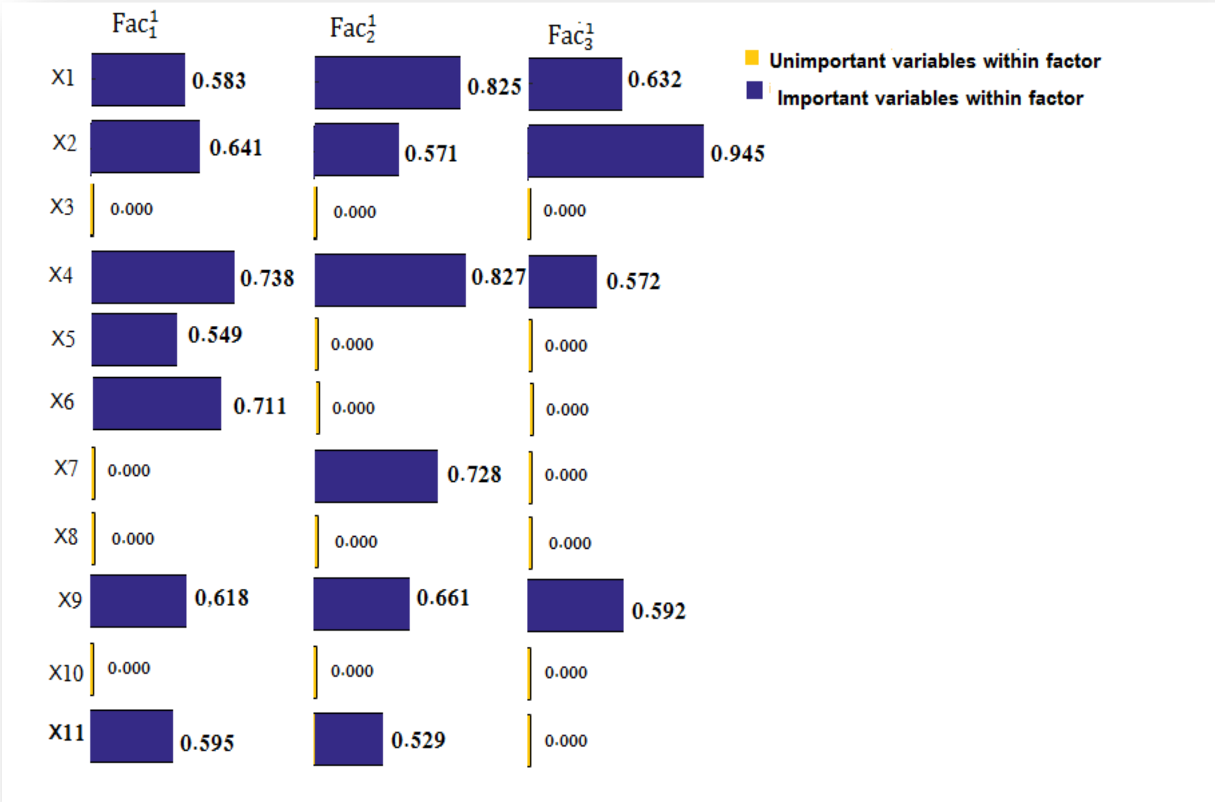

Figure -4- Show the factor loading of original variable with important second –order factor by proposed method (Lasso high-order factor analysis)

From the results listed in table 2 ,we observed there are three factors non-zero eigenvalues, therefore, the number of important factors are three, these three factor can explain 91.001% from total variance. From above figure displayed three important factor, we can explain the loading factor of original variable with three important factors are following : In first second-order factor can explain 32.784% from total variance , and this factor have seven non-zero loading of the original variables($X_1, X_2, X_4, X_5, X_6, X_9, X_{11}$).Also, this factor have four zero exactly loading of the original variables(X_3, X_7, X_8, X_{10}),these variables unimportant for analysis. In second factor of the second-order factor can explain 30.563% from total variance , and this factor have four non-zero loading of the original variables($X_1, X_2, X_4, X_7, X_9, X_{11}$).Also, this factor have four zero exactly loading of the original variables($X_3, X_5, X_6, X_8, X_{10}$),these variables unimportant for analysis. In third second-order factor can explain 27.654% from total variance , and this factor have seven non-zero loading of the original variables($X_1, X_2, X_4, X_9,$).Also, this factor have seven zero exactly loading of the original variables($X_3, X_5, X_6, X_{7,} X_{8,} X_{10,} X_{11,}$),these variables unimportant for analysis.

Conclusions and recommendations

Conclusions

important factor Selection with high-order factor analysis, regularization methods like Lasso can be used to efficiently choose important factor . In order to reduce dimensionality and concentrate on the most significant factor , which phase is essential. Also, Regularization reduces the possibility of overfitting, especially with high-dimensional datasets. Constraining the model guarantees the generalizability of the results. Automatic Factor Selection: By setting inconsequential coefficients to zero, regularization makes it easier for major factors to be automatically selected. This automaticity lessens the subjectivity in factor selection and streamlines the analytical process. Comparative Advantage: The regularization technique shows that when comparing factor analyses of the first ,second to the high orders, the high order frequently produces a larger explained variance with fewer significant components. This demonstrates how well hierarchical modeling captures intricate interactions. The two-stage high-order factor analysis approach is greatly improved with the addition of regularization techniques. While addressing problems like overfitting, it enhances robustness, high model interpretability, and variable selection within significant factors. These benefits highlight regularization's importance as a necessary element of contemporary analytical frameworks. We find that our proposed method demonstrated a greater ability to explain significant factors compared to the classical method, as evidenced by the explained variance. Our proposed method was able to explain a higher percentage of variance and was also effective with high-order factor analysis. Additionally, our proposed method does not rely on ranking when selecting significant factors or when choosing the original variables within those important factors. We note that the variable (Living conditions) has the greatest impact on the phenomenon of divorce within the first and second factors, as it achieved the highest factor loadings within these two factors. Also, the variables (Changes in values)(Increased independence of women.) have the greatest impact on the phenomenon of divorce within the first and second factors,

Recommendations

Combine a good regularization methods To determine which regularization approach yields the best results for a given dataset, investigate the employment of various regularization techniques (such as Elastic Net and Ridge) in conjunction with Lasso. Also, To assess the impact of regularization parameter selection on the selection of important factors and the overall performance of the model, via sensitivity studies. In high-dimensional phenomena, we recommend using our proposed method due to its advantages in identifying the appropriate level of analysis that achieves the highest explained variance. Additionally, the combination of the Lasso technique with high-order factor analysis provides a new benchmark for selecting significant factors as well as important variables within the chosen factors. We recommend that civil society organizations hold awareness seminars to reduce this phenomenon by providing a wider space for women's independence and supporting independent living away from family and relatives' interference.

Reference

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research. Guilford Press.Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate behavioral research*, *25*(1), 33-39.

Cieciuch, J., Davidov, E., Vecchione, M., & Schwartz, S. H. (2014). A hierarchical structure of basic human values in a third-order confirmatory factor analysis. Swiss Journal of Psychology.

Costello, A. B., & Osborne, J. (2019). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.

Cudeck, R. (2000). Exploratory factor analysis. In *Handbook of applied multivariate statistics and mathematical modeling* (pp. 265-296). Academic Press.

Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics. Sage Publications.

Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comments on Improved Practice. Educational and Psychological Measurement, 66(3), 393-416.

Jolliffe, I. T. (2002), Principal Component Analysis (2nd ed.). Springer.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement, 20(1), 141-151.

Kim, J. O., & Mueller, C. W. (1978). Factor Analysis: Statistical Methods and Practical Issues. Sage Publications.

Okon, Jan(1974). Factor Analysis, Trans, Russian, statistica Moscow.

Taha Hussein Ali, and Fadhel Hamid Hadi(2010). “Using Simulation to Determine the Optimal Degree in Principal Component Analysis of Higher Degrees.” Journal of Al-Qadisiyah for Computer Science and Mathematics 2.1.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society: Series B (Statistical Methodology), 58(1), 267-288.

Toczydlowska, D., Peters, G. W., Fung, M. C., & Shevchenko, P. V. (2017). Stochastic period and cohort effect state-space mortality models incorporating demographic factors via probabilistic robust principal components. *Risks*, *5*(3), 42.