International Journal of Research Methodology (IJRM)

Volume 02, Issue 10, October 2024 **ijrm.humanjournals.com** ISSN: 2584-1793

Exploring Factor Scores, Structure Coefficient, and Communality Coefficient

Lionel K. Asare

Department of Sociology, Anthropology, Human Development and Family Science,

University of Louisiana, Lafayette, LA 70504 U.S.

Received: 2024-10-04 Revised: 2024-10-17 Accepted: 2024-10-21

ABSTRACT

Factor analysis is a statistical tool that is utilized to summarize the relationships among variables in a concise manner to help in conceptualization. It is regarded as one of the most powerful methods for reducing variable complexity to greater simplicity. This research report aims at explaining the meaning and usage of three useful factor analytic statistics: factor scores, structure coefficient, and communality coefficient in a simplistic way to enable researchers to correctly identify and interpret these concepts. A better understanding of these concepts will help researchers make sense of factor analytic results.

Keywords*:* factor analysis, effect size, statistics, parametric analysis

INTRODUCTION

Factor Analysis

Exploratory factor analysis (EFA) has been used as an analytical tool in educational research for a variety of purposes like the reduction of large number of items from a questionnaire or survey instrument to a smaller number of components (Distefano, Zhu &Mindrila,2009). This method helps in uncovering latent dimensions underlying a data set, or examining which items have the strongest association with a given factor. After identifying the number of factors using EFA, applied researcher use the information for ensuing analyses (Gorsuch, 1983).In other to use EFA information in any follow-up studies, applied researchers must create scores to represent each individual's placement on the factor(s) identified from the EFA. These factor scores may then be used to investigate the research questions of interest.

The purpose of this paper is to facilitate an understanding of the relationships between these 3 factor analytic statistics: Factor scores, structure coefficient, and communality coefficient. This paper will utilize a heuristic data set from Holzinger and Swineford (1993) to demonstrate these relationships. The paper will further illustrate the logic behind the computation and interpretation of these factor analytic statistics in a factor analysis interpretation by describing ways in which a researcher may create factor scores following an EFA, and how to also compute structure coefficient, and communality coefficients. Finally, this paper will discuss the advantages and disadvantages among the methods of computing factor scores. Factor analysis has been touted as one of the most effective methods for reducing variable complexity to greater simplicity. It has also been regarded as the furthest logical development and "reigning queen" of the correlational methods (Thompson, 2004).

Factor Analysis and Other Parametric Analysis

Factor analysis is part of the classical parametric analysis which implies that it is correlational and applies weight to observed variables to create synthetic variables that becomes the focus of all analysis. Factor analysis, like all parametric analysis yield variance-accounted for effect sizes that are analogous to r^2 (Thompson, 1991). However, in spite of the usefulness of factor analysis in applied research, not all commentaries on factor analysis have been welcoming in the field of research. Notable among these is the misuse of factor analysis and the confusing language associated with the entire method. The confusing language associated with factor analysis stems from the fact that it belongs to the parametric analytic family and shares the same characteristic with them: all correlational in nature, apply weights to observed variables to create synthetic variables and that these synthetic variables becomes the focus of all analyses and finally, it yields a variance accounted for effect sizes that are analogous to r^2 in multiple regression (Sherry & Henson, 2005).

The implication for this is that, factor analysis call the same system of weights equations in regression, factors in factor analysis, functions or rules in discriminant analysis and canonical correlational analysis. In addition, the weights are utilized in factor analysis.

These weights are termed beta weights in regression, pattern coefficient in factor analysis, and standardized function coefficient in discriminant analysis or canonical correlational analysis. Factor analysis also yield's synthetic scores that are analogous to *yhat* in regression, discriminant scores in discriminant analysis, and canonical function score in canonical correlational analysis. These scores are what we call factor scores and that is what is going to be one of the major focus of discussion in this paper.

The goal of this paper is to demonstrate that statistical methodologies used in factor analysis is not that different from the other parametric methods mentioned above. Thompson (2004), commenting of these, mentioned that the different names in different analysis are just to confuse graduate students. This paper will try to clarify this confusion by explaining the meaning and use of these three important factor analytic statistics: *factor scores, factor structure coefficients, and communality coefficient.* An understanding of these terms will help educational researchers to make sense of factor analytic results. As already stated, part of a data set from Holzinger and Swineford (1939) will be used to demonstrate the computation and interpretation of these statistics.

Factor Scores

The initial step in factor analysis is the creation of a matrix of association coefficient from a raw data matrix. There are a variety of matrix of association that could be utilized to compute this matrix of association of the observed variables. Some of the most popular matrix of associations are Pearson correlation, variance –covariance matrix, and a spearman rho matrix.

As already indicated, exploratory factor analysis allows researchers to compute scores for the individuals in the analysis on the extracted factors. These scores (factor scores) can therefore be used in a wide variety of subsequent statistical analysis. Factor scores can simply be defined as numerical values that indicates an individual's relative standing on a latent factor. In other words, factor score estimates are not perfectly correlated with the factors themselves, and depending on the nature of a particular data set these estimates may in fact only be moderately correlated with the factors.

These individual scores extracted from EFA, according to Grice (2001) can be correlated with measures of different constructs to help clarify the nature of the factors or they can be entered as predictor variables in multiple regression analyses or as dependent variable in analysis of variance. It should be noted that factor scores computed as simple sum scores in scale development process are often referred to as composite or total scores. The choice made regarding how factor scores are computed can significantly affect the quality of the factor scores (Grice, 2001). When computing factor scores, factor pattern matrix, also called weight matrix is applied to the observed or measured variables. This process is analogous to *yhat* scores in multiple regression. The factor pattern coefficients are also analogous to the regression weights called beta weights. One important thing however with factor analysis is that, if the pattern and structure coefficients are not identical, the factor pattern coefficients must be utilized to compute factor scores (Distefano, Zhu, & Mindrila, 2009).

Gorsuch (1983) outlined four systematic procedures needed in computing factor scores. To begin with, Gorsuch (1983), posited that there should be a high correlation between the common factor scores and the factors that you are attempting to measure. Secondly, the common factor scores should be unbiased estimates of the true factor scores. In addition, when factors are orthogonal, the factor score estimates from one factor should have zero correlations with all the other factors. Finally, if the factors are orthogonal, the common factor scores should correlate with each other. The strategies for calculating factor scores that will be discussed in this paper includes the regression method, the Bartlett method, the Anderson-Rubin method, and finally, the Thompson method.

Regression Method

Regression factor scores predict the location of each individual on the factor or component. As the name implies, independent variables in the regression equation are the standardized observed values of the items in the estimated factors or components. The matrix formula for calculating factor scores using the regression method is given as follows:

$$Z_{N*V}\, W_{V*F} = F_{N*F}$$

Where Z_{N*V} is the *Z-score* matrix of *N* people on the *V* variables, W_{V*F} is the weight matrix applied to the *V* variables to obtain the *F* factor scores for each of the items of measure. F_{N*F} is the factor score matrix of *N* individuals on the *F* factors. The weight matrix W_{V*F} is itself a product of the R_{V*V}^{-1}, the inverse of the *V* variable by *V* variable correlation matrix, and P_{V*F}, which is the factor pattern coefficient matrix. Therefore the matrix formula for factor scores for the regression method can be rewritten as: $Z_{N*V}\, R_{V*V}^{-}\, P_{V*F}, = F_{N*F}.$

The regression method remains a popular choice for calculating factor scores because of researchers' widespread familiarity with multiple regression techniques. Regression factor scores can readily be computed as an option within SPSS using the FACTOR subcommand: SAVE=REG (ALLFSCORE). The factor score weight matrix labeled by SPSS the Factor Score Coefficient Matrix, can be requested as an optional out. Table 1 presents this matrix from the data utilized for this paper. Lastly, each person's *Zscore* on the six measured variables are multiplied by the weight matrix values for both factors, and these products are then summed, thus yielding two factor scores for each person. For factor 1, per the table, the compute statement that accomplishes this would be:

COMPUTE FSHARD1= (.39949*ZT10) + (.45189*ZT12) + (.38404*ZT13) + (.-10139*ZT14) + (-.10055*ZT15) + (.11149*ZT17).

So for person number from the data set of 301 people, this person's synthetic variable score on factor 1 would be computed as

= (.39949*-.72930) + (.45189*.22015) + (.38404*.97804) + (.-10139*-.44778) + (-.10055*-.51880) + (.11149*-.45415).

= .-29135+.09948+.37561+.04540+.05217+-.05063= **.230674**

For Factor II, the COMPUTE statement would be: = (-.01362*ZT10) + (-.10183*ZT12) + (-.04562*ZT13) + (.50114*ZT14) + (.48359*ZT15) + (.35269*ZT17).

The Bartlett's Method

Another way of computing factor scores is the use of the Bartlett method. In contrast to the regression method, only the shared (i.e., common) factors have an impact on factor scores. The sum of squared components for the "error" factors (i.e., unique factors) across the set of variables is minimized. Therefore the resulting factor scores are highly correlated with their corresponding factor and not with other factors. However, the estimated factor scores between different factors may still correlate. Bartlett factor scores are computed by finding the product of the row vector of observed variables and the inverse of the diagonal matrix of variances of the unique factor scores and the factor pattern matrix of loadings. Resulting values are then multiplied by the inverse of the matrix product of the matrices of factor loadings and the inverse of the diagonal matrix of variances of the unique factor scores (DiStefano, Zhu, & Mindrila, 2009). One advantage of Bartlett factor scores is that it produces unbiased estimates of the true factor scores (Distefano, Zhu, &Mindrila, 2009). This is because Bartlett scores are produced by using maximum likelihood estimates – a statistical procedure which produces estimates that are the most likely to represent the true factor scores.

The Anderson-Rubin method

This method was proposed by Anderson and Rubin (1956). It has a lot of similarity with the Bartlett method except that Anderson and Rubin added a requirement for the factor scores to be uncorrelated. That is, the least squares formula is adjusted to produce factor scores that are uncorrelated with other factors and also with each other. The Anderson-Rubin equation is more complex than Bartlett's. However, the factor estimates generated by this method has correlations that form an identity matrix. The Anderson-Rubin estimates, like the other two already discussed can be automatically generated in SPSS by selecting the Anderson and Rubin option in the Factor Analysis: Factor Scores window.

Thompson Method

Thompson (1993) came up with another method for computing factor scores which is very different from the methods discussed so far. One general similarity with the methods already discussed is that, the generated factor scores are in *Zscore* format. Implying that each set of factor score for any given factor will have a mean of zero and a standard deviation of one. Wells (1999) posited that when factor scores are in *Zscore* forms, it does not allow the researcher to compare the mean factor score on any given factor with the factor score means on other factors for the same data set. That explains one advantage that the Thompson method has over the other methods. It yields a standardized, noncentered factor score that permits the comparison of factor score means across factors. Therefore this method yield factor scores with a standard deviation of one, but a non-zero mean.

The calculation of factor scores utilizing the Thompson method is given as follows: To begin with, the variables are converted to *Zscore* form using the SAVE option in DESCRIPTIVES procedure within SPSS. In the second step of this method, the original variable means are added back onto the Z-scores using COMPUTE statements. For example, the measured variable, *T10*, would be re expressed as

COMPUTE TT10=ZT10 + 96.28

The variable TT10 would have its original mean, but the standard deviation of 1. In the third step of this method, factor scores are obtained by multiplying these standardize but non-centered scores by the values in the factor score matrix, W_{V*F} , as previously illustrated. These steps are performed separately for each factor score composite. The Thompson factor score for factor1 would be calculated as:

International Journal of Research Methodology (IJRM)
Volume 02, Issue 10, October 2024 ijrm.humanjournals.com ISSN: 2584-1793

COMPUTE FSBT1 (.39949*TT10) + (.45189*TT12) + (.38404*TT13) + (.-10139*TT14) + (-.10055*TT15) + (.11149*TT17)

Table 1. Varimax-rotated Pattern/Structure Coefficients

	Factor		
Variable	*I*	*II*	h^2
T10	.13724	.76679	.60680
T12	.00910	.83268	.69344
T13	.07693	.72407	.53020
T14	.80670	.00505	.65079
T15	.77737	-.00036	.60430
T17	.64109	.35701	.53845

The communality coefficient, h^2 can be computed by squaring the pattern/structure coefficients, and adding all these squared values for a given measured variable (e.g., T10, $.76679^2 + .13724^2 = 58.797\% + 1.883\% = 60.680\%$.

Relationships among Factor Scores

It can be seen from the table 2 that if and only if factor extraction is by principal components methods, all factor scores obtained by the various methods will be identical, except for the factor scores generated via Thompson methods. It is important to understand that although the means of the factor scores generated by the four methods are not equal, when the correlations among the factor scores are examined, all the factor scores correlate perfectly with all the other factor scores on the same factor. This occurs only if the principal components is the factor extraction method.

Structure and Communality Coefficient

In parametric analysis, structure coefficient are called such because of how they inform researchers about the structure or makeup of the effect represented by the synthetic variable *yhat*. It is a simple bivariate correlation between an observed variable and a synthetic variable. It should be noted that, because they are bivariate correlations, they are not affected by collinearity between the predictors. They simply throw an important light on the importance of predictors. In factor analysis, when the factors are not perfectly uncorrelated, as typical in an oblique rotation, the factor pattern matrix and factor structure matrix will not be identical, therefore both must be interpreted. This is analogous to the multiple regression analysis in the sense that, sometimes the beta weights and the structure coefficients are equal but usually both must be interpreted. Elements within the factor structure matrix, S_{F*F} are called structure coefficient.

Communality coefficient on the other hand is explained as the proportion of variance in the observed variable that is reproduced in the extracted factors. The communality coefficient for a given observed variable is generated by the sum of the squared structure coefficient across each row of the factor structure matrix, as noted previously. Communality coefficient is denoted by the symbol h^2 .Because h^2 a squared metric statistic, it can be operationalized as the R^2 between a given measured variable and the two factor scores.

Table 2

			Factor Score Computation Method		
Person	*Regression*[a]	*Bartlett*	*Anderson-Rubin*	*Regression*[b]	*Thompson*
1	.23066	.23066	.23066	.23	137.04
2	1.21497	1.21497	1.21497	1.22	138.02
3	-1.48309	-1.48309	-1.48309	-1.48	135.32
4	-.72601	-.72601	-.72601	-.72	136.08
5	.30378	.30378	.30378	.30	

[a] *These regression factor scores were computed using the SPSS "save" command.*

Conclusion and Implication for Applied Researchers

In conclusion, the purpose of this paper was to simplify and explain factor score, structure coefficient, and communality coefficient in a way that upcoming researchers can understand and also to overcome the confusion that underlies these important concepts in factor analysis. It is often time difficult for readers to understand these concepts because of the confusing practice of using different terminologies in different analyses to name the same concept. Factor analysis is an important statistical tool that, when properly utilized can help researchers explore complex network of interrelationships among variables. This paper helped to reiterate the fact that factor scores are simply latent variables, or weighted combinations of observed scores. That is, this paper has established the fact that factor analysis helps in uncovering latent dimensions underlying a data set, or examining which items have the strongest association with a given factor. Applied researchers can use this invaluable information for future research analysis utilizing this statistical technique.

Using factor analysis can help to explore a complex network of relationship among variables. Factor scores in EFA analysis allows the researcher to compute scores for the individuals in the analysis on the extracted factors. These scores (factor scores) are subsequently used in in a wide variety of statistical analysis. The relationships among the various methods for computing factor score goes to buttress the fact that all parametric analysis are similar as could be seen in the demonstrations above. Furthermore emphasizing the importance of interpreting structure coefficients and beta weight when the two are not equal

Appendix: Syntax for Analysis

SUBTITLE '1. show what structure coefficients are $$$$'.

CORRELATIONS VARIABLES T10 T12 T13 T14 T15 T17 WITH FSCORES1 FSCORES2.

subtitle '2a. show what communality coefficients are##'.

regression

variables=T10 T12 T13 T14 T15 T17 FSCORE1 FSCORE2/DEPENDENT= T10/enter fscore1 fscore2 .

subtitle '2b. show what communality coefficients are##' .

regression

variables= T10 T12 T13 T14 T15 T17 FSCORE1 FSCORE2/ dependent= T12/enter fscore1 fscore2 .

subtitle '2c . show what communality coefficients are##' .

regression

variables=T10 T12 T13 T14 T15 T17 FSCORE1 FSCORE2/dependent=T13/enter fscore1 fscore2 .

subtitle '2d. show what communality coefficients are##' .

regression

variables=T10 T12 T13 T14 T15 T17 FSCORE1 FSCORE2/dependent=T14/enter fscore1 fscore2 .

subtitle '2e. show what communality coefficients are##' .

regression

variables=T10 T12 T13 T14 T15 T17 FSCORE1 FSCORE2/dependent=T15/enter fscore1 fscore2 .

subtitle '2f. show what communality coefficients are##' .

regression

variables=T10 T12 T13 T14 T15 T17 FSCORE1 FSCORE2/dependent=T17/enter fscore1 fscore2 .

SUBTITLE '3a. Factor scores BARTLETT method ###'.

FACTOR VARIABLES=T10 T12 T13 T14 T15 T17/CRITERIA=FACTORS (2)

/EXTRACTION=PC/ROTATION=VARIMAX/SAVE=BART (ALL FSCOR).

VARIABLE LABELS FSCOR1 'SPEED bart'

FSCOR2 'MEMORY bart'.

SUBTITLE '3b. Factor scores ANDERSON-RUBIN ####'.

FACTOR VARIABLES=T10 T12 T13 T14 T15 T17/CRITERIA=FACTORS (2)

/EXTRACTION=PC/ROTATION=VARIMAX/SAVE=AR (ALL FSCR).

VARIABLE LABELS FSCR1 'SPEED ar'

FSCR2 'MEMORY ar' .

SUBTITLE '4a. compute z-score ****' .

DESCRIPTIVES VARIABLES= T10 TO T17/save.

print formats zt10 to zt17 (F8.5) .

list variables=zt10 to zt17/cases=25 .

SUBTITLE '4b. Prove z-scores are z-scores ****'.

DESCRIPTIVES VARIABLES=ZT10 TO ZT17.

SUBTITLE '4c. compute regression factor scores hard way ****' .

COMPUTE FSHARD1= (.39949*ZT10)+(.45189*ZT12)+(.38404*ZT13)+(-.10139*ZT14)+(-.10055*)+(.11149*ZT17).

COMPUTE FSHARD2=(-.01362*ZT10)+(-.10183*ZT12)+(-.04562*ZT13)+(.50114*ZT14)+(.48359*ZT15)+(.35269*ZT17).

VARIABLE LABELS FSHARD1 'SPEED hard'

FSHARD2 'MEMORY hard'.

SUBTITLE '5a. compute Thompson factor scores @@@@' .

COMPUTE TT10=ZT10+96.28.

COMPUTE TT12=ZT12+110.54.

COMPUTE TT13=ZT13+193.47.

COMPUTE TT14=ZT14+175.15.

COMPUTE TT15=ZT15+90.01.

COMPUTE TT17=ZT17+8.23.

COMPUTE FSBT1=(.39949*ZT10)+(.45189*ZT12)+(.38404*ZT13)+(-.10139*ZT14)+(-.10055*ZT15)+(.11149*ZT17).

COMPUTE FSBT2=(-.01362*ZT10)+(-.10183*ZT12)+(-.04562*ZT13)+(.50114*ZT14)+(.48359*ZT15)+(.35269*ZT17).

VARIABLE LABELS FSBT1 'SPEED thompson'

FSBT2 'MEMORY thompson'.

SUBTITLE '6. Show factor score relationships &&&&'.

LIST VARIABLES=FSCORE1 FSCR1 FSHARD1 FSBT1/CASES=25.

DESCRIPTIVES VARIABLES=FSCORE1 TO FSCR2 FSHARD1 TO FSHARD2 FSBT1 TO FSBT2.

CORRELATIONS VARIABLES= FSCORE1 TO FSCR2 FSHARD1 TO FSHARD2 FSBT1 TO FSBT2.

REFERENCES

1. DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation,14* (20),1-11
2. Gorsuch, R., L. (1983). *Factor analysis* (2^{nd} ed.). Hillsdale, NJ: Erlbaum.
3. Grice, J., W. (2001). Computing and evaluating factor score. *Psychological Methods, 6*(4), 430-450.
4. Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (No. 48). Chicago: University of Chicago.
5. Sherry, A., & Henson, R., K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment, 84*(1), 37-48.
6. Thompson, B. (2004).*Exploratory and Confirmatory Factor Analysis* Washington, DC: American Psychological Association.
7. Wells, R., D. (1999). Factor scores and factor structure and communality coefficients. *Advances in Social Sciences, 5,* 123-138

How to cite this article:
Lionel K. Asare. Ijrm.Human, 2024; Vol. 02 (10): 1-7.
Conflict of Interest Statement: All authors have nothing else to disclose.