Article: RT1006

Biotica Research Today
Vol 4:6
2022
432
434

How to cite this article?
Chittora *et al.*, 2022. Factor Analysis: A Data Reduction Technique. Biotica Research Today 4(6):432-434.

Factor Analysis: A Data Reduction Technique

Vaibhav Chittora[1*], Heerendra Prasad[1], Prashant Vasishth[2] and Mohit Sharma[2]

[1]Dr. YSPUHF, Nauni, Solan, Himachal Pradesh (173 230), India

[2]ICAR-Indian Agricultural Research Institute, Pusa, New Delhi, Delhi (110 012), India

Open Access

Corresponding Author
Vaibhav Chittora
e-mail: vchittora97@gmail.com

Keywords
Correlation, Matrix, MLR, VIF

Abstract

In many studies we observe large number of variables and in these variables many of them gives common information that is why it is not worthy to study all the variables as it complicates analysis and interpretation. There it comes the use of data reduction techniques. These techniques extract some most influencing variables from the large variables. Factor analysis (FA) is a multivariate technique that is used to describe the relationships between different variables under study (observable variables) with new variables called factors, where the number of factors is less than the number of original variables. FA works efficiently and produces fewer factors to describe the relationship if the variables under study are highly correlated. For instance, if all of the variables in one group are highly correlated among themselves and have little correlation with the variables in the remaining groups, each group can represent a factor. FA is considered an extension of principal component analysis since the ultimate objective for both techniques is a data reduction.

Introduction

A data reduction or dimension reduction technique is designed to represent a wide range of attributes on a smaller number of dimensions.

"A statistical approach that can be used to analyze interrelationship among a large number of variables and explain these variables in terms of their common underlying dimension (factor)".

• Factor analysis is a general name denoting a class of procedures primarily used for data reduction and summarization.

• Variables are not classified as either dependent or independent. Instead, the whole set of interdependent relationships among variables is examined in order to define a set of common dimensions called Factors.

Types of Factor Analysis

Exploratory Factor Analysis

• When the dimensions/ factors are theoretically unknown.

• Exploratory Factor Analysis (EFA) is a statistical approach to determining the correlation among the variables in a dataset. This type of analysis provides a factor structure (a grouping of variables based on strong correlations).

Confirmatory Factor Analysis

• When researcher has preconceived thoughts about the actual structure of data based on theoretical support or prior research.

• Researcher may wish to test hypothesis involving issues as which variables should be grouped together on a factor (Anderson, 1958).

Article History
Received on: 31st May 2022
Revised on: 13th June 2022
Accepted on: 14th June 2022

E-mail: *bioticapublications@gmail.com*

Purpose of Factor Analysis

• To identify underlying dimensions called Factors, that explain the correlations among a set of variables.

• To reduce the number of variables (data reduction).

• To identify a new, smaller set of uncorrelated variables to replace the original set of correlated variables for subsequent analysis such as Regression or Discriminant Analysis.

Assumptions

• Models are usually based on linear relationships.

• Models assume that the data collected are continuous in nature.

• Multicollinearity in the data is desirable because the objective is to identify interrelated set of variables.

• The data should be amenable for factor analysis. It should not be such that a variable is only correlated with itself and no correlation exists with any other variables. This is like an Identity Matrix. Factor analysis cannot be done on such data (Agarwal, 2003).

• Sample size: Minimum numbers of variable for FA is 5 cases per variable *e.g.*, 20 variables should have > 100 cases (1:5) ideal condition.

Partitioning the Variance in Factor Analysis

Since the goal of factor analysis is to model the interrelationships among items, we focus primarily on the variance and covariance rather than the mean.

• **Common variance**: Common variance is the amount of variance that is shared among a set of items. Items that are highly correlated will share a lot of variance.

• **Unique variance**: Unique variance is any portion of variance that's not common.

There are two types of Unique variance.

• **Specific variance:** Specific variance is variance that is specific to a particular item.

• **Error variance:** Error variance comes from errors of measurement and basically anything unexplained by common or specific variance.

The total variance is made up to common variance and unique variance.

The goal of a factor analysis is to reduce the number of variables to explain and to interpret the results. This can be accomplished in two steps:

1. Factor extraction

2. Factor rotation

Factor extraction involves making a choice about the type of model as well the number of factors to extract.

Factor rotation comes after the factors are extracted, with the goal of achieving simple structure in order to improve interpretability.

Extracting Factors

There are two approaches to factor extraction which stems from different approaches to variance partitioning:

a) Principal Components Analysis

b) Common Factor Analysis

a) Principal Components Analysis

• PCA is a dimension-reduction tool that can be used to reduce large set of variables to a small set that still contains most of the information in the large set.

• Principal Components Analysis or PCA makes the assumption that there is no unique variance; the total variance is equal to common variance.

• Additionally, if the total variance is 1, then the common variance is equal to the communality (Johnson and Wichern, 2009).

b) Common Factor Analysis

• Common Factor Analysis assumes that total variance can be partitioned into common and unique variance.

• The other main difference between PCA and factor analysis lies in the goal of our analysis.

• If our goal is to simply reduce our variable list down into a linear combination of smaller components then PCA is the way to go.

• However, if you believe there is some latent construct that defines the interrelationship among items, then factor analysis may be more appropriate.

• PCA may be used to decompose a correlation matrix or a covariance matrix.

• Data must be tested for linearity and normality before continuing with PCA.

• If the data do not substantially violate linear relationships and normal distributions, use of PCA on a correlation matrix is an appropriate and a valuable data reduction technique.

•Otherwise use of PCA of covariance matrix is appropriate.

Conclusion

Data reduction techniques help in decrease in dimension of the data. It makes data simple to represent and runs through advanced analytical algorithm. It also helps in making data easy to handle, analyse and interpret in multiple regression analysis as it makes all the regressor variable independent.

References

Agarwal, B.L., 2003. Programmed Statistics. New Age International Publishers, New Delhi, p. 630.

Anderson, T.W., 1958. An Introduction to Multivariate Analysis, John Wiley, New York, p. 569.

Johnson, R.A., Wichern D.W., 2009. Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey, 5th edition, p. 481.