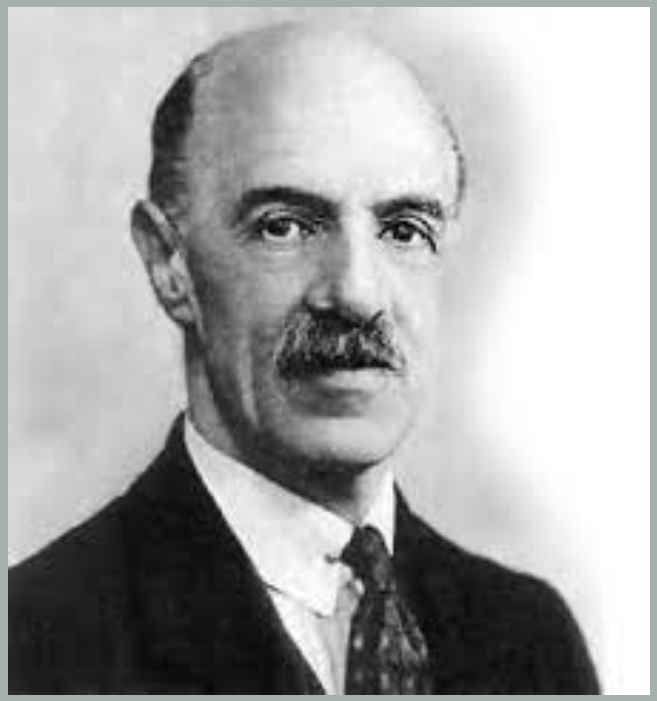
ICCPP-STATISTICS

- Factor Analysis

Vishal Lohchab

Scientific Assistant of
Prof. Dr. Hans-Werner Gessmann
Director ICCPP International



Charles Spearman (1863-1945)

Factor Analysis

Definition

- Factor analysis is a statistical method used to describe factors among observed, correlated variables in terms of a potentially lower number of unobserved factors.

Factor Analysis

- Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors.
- This technique extracts maximum common variance from all variables and puts them into a common score. As an index of all variables, we can use this score for further analysis.

Factor Analysis

- A "factor" is a set of observed variables that have similar response patterns; they are associated with a hidden variable (called a confounding variable) that isn't directly measured.
- Factors are listed according to factor loadings, or how much variation in the data they can explain.

The two types of Factor Analysis

1. Exploratory factor analysis is used if you don't know the data structure or how many dimensions are in a set of variables.
2. Confirmatory factor analysis is used for verification as long as you have a specific idea about what the data structure is or how many dimensions are in a set of variables.

Confirmatory Factor Analysis

- Confirmatory Factor Analysis allows you to figure out if a relationship between a set of observed variables and their underlying constructs exists.
- It is similar to Exploratory Factor Analysis. The main difference between the two is:

  If you want to explore patterns, use EFA.
  If you want to perform hypothesis testing, use CFA.

Confirmatory Factor Analysis

- EFA provides information about the optimal number of factors required to represent the data set. With CFA you can specify the number of factors required.
- For example, CFA can answer questions like "Does my ten questions survey measure one or more specific factors?".

Exploratory Factor Analysis

- EFA is used to find the underlying structure of a large set of variables. It reduces data to a much smaller set of summary factors.
- EFA is almost identical to CFA.

Exploratory Factor Analysis

- Similarities are:

  Assess the internal reliability of a measure.

  Examine factors or theoretical constructs represented by item sets. They assume the factors aren't correlated.

  Investigate quality for individual items.

Exploratory Factor Analysis

- There are, however, some differences, mostly concerning how factors are treated/used.
- EFA is basically a data-driven approach, allowing all items to load on all factors, while with CFA you must specify which factors to load.

Use of EFA and CFA

- EFA is a good choice if you don't have any idea about what common factors might exists. EFA can generate a large number of possible models for your data, something that may not be possible if a researcher has to specify factors.
- If you do have an idea about what the models look like, and you want to test your hypotheses about the data structure, CFA is a better approach.

Factor loading

- Factor loading is basically the correlation coefficient for the variable and factor. Factor loading shows the variance explained by the variable on that particular factor.
- In the SEM approach, as a rule of thumb, 0.7 or higher factor loading represents that the factor extracts sufficient variance from that variable.

Eigenvalues

- Eigenvalues is also called characteristic roots. Eigenvalues shows variance explained by that particular factor out of the total variance.
- From the commonality column, we can know how much variance is explained by the first factor out of the total variance.

Factor score

- The factor score is also called the component score. This score is of all row and columns, which can be used as an index of all variables and can be used for further analysis.
- We can standardize this score by multiplying a common term. With this factor score, whatever analysis we will do, we will assume that all variables will behave as factor scores and will move.

Criteria for determining the number of factors

- According to the Kaiser Criterion, Eigenvalues is a good criteria for determining a factor.
- If Eigenvalues is greater than one, we should consider that a factor and if Eigenvalues is less than one, then we should not consider that a factor. According to the variance extraction rule, it should be more than 0.7. If variance is less than 0.7, then we should not consider that a factor.

Rotation method

- Rotation method makes it more reliable to understand the output.
- Eigenvalues do not affect the rotation method, but the rotation method affects the Eigenvalues or percentage of variance extracted.
- Rotations minimize the complexity of the factor loadings to make the structure simpler to interpret.

Rotation method

- There are a number of rotation methods available:
  (1) No rotation method,
  (2) Varimax rotation method,
  (3) Quartimax rotation method,
  (4) Direct oblimin rotation method,
  (5) Promax rotation method.

- Each of these can be easily selected in SPSS, and we can compare our variance explained by those particular methods.

Performing Factor Analysis

Factor Analysis is an extremely complex mathematical procedure and is performed with statistical software. There are many software's to perform FA. You have to follow the instructions of software you want to use.

- Name of some Softwares are:
  1. Stata
  2. Minitab
  3. SPSS

Performing Factor Analysis

- Use Kaiser-Meyer-Olkin test to see if your data is suitable for Factor Analysis.
- The statistic is a measure of the proportion of variance among variables that might be common variance. The lower the proportion, the more suited your data is to Factor Analysis.

Performing Factor Analysis

KMO returns values between 0 and 1. A rule of thumb for interpreting the statistic:

- KMO values between 0.8 and 1 indicate the sampling is adequate.
- KMO values less than 0.6 indicate the sampling is not adequate and that remedial action should be taken.
  Some authors put this value at 0.5, so use your own judgment for values between 0.5 and 0.6.

Performing Factor Analysis

- KMO Values close to zero means that there are large partial correlations compared to the sum of correlations.
  In other words, there are widespread correlations which are a large problem for factor analysis.

References

https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/factor-analysis/, date 22.11.21, 16:30 h MET

https://www.definitions.net/definition/factor+analysis, date 22.11.21, 16:00 h MET

https://www.statisticshowto.com/factor-analysis/, date 22.11.21, 17:00 h MET

Dodge, Y. (2008). The Concise Encyclopedia of Statistics. Springer.

Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate analysis. Washington, DC: American Psychological Association.