MIS 652-E: Multivariate Data Analysis I

School of Business at Stevens Institute of Technology

Fall 2017

Chapter 3 Assignment

Name: Balaji Katakam

Date: 10/29/2017

Q1) What are the difference between the objectives of data summarization and data reduction?

Solution 1) The essential contrast between the targets of data summarization and data reduction relies on a definitive research question. In data summarization a definitive research question might be to better understand the interrelationship among the factors. This might be refined by consolidating an extensive number of respondents into fewer particularly unique gatherings with Q-sort factor investigation. All the more frequently data summarization is connected to factors in R-sort factor examination to recognize the measurements that are idle inside a dataset. Data summarization makes the distinguishing proof and understanding of these fundamental measurements or variables a definitive research question.

Data reduction depends on the ID of the measurements too, however makes utilization of the disclosure of the things that include the measurements to lessen the data to less factors that speak to the inert measurements. This is proficient by either the utilization of surrogate factors, summated scales, or factor scores. Once the data has been decreased to the less number of factors assist examination may end up plainly less demanding to perform and decipher.

Q2) How can factor analysis help the researcher improve the results of other multivariate techniques?

Solution 2) Factor analysis gives coordinate knowledge into the interrelationships among factors or respondents through its information condensing point of view. This gives the analyst an unmistakable picture of which factors are very connected and will act in show in different analysis. The outline may likewise prompt a superior comprehension of the inert measurements basic an exploration question that is at last being replied with another technique. From an information decrease point of view, the factor analysis comes about permit the development of surrogate or summated factors to speak to the first factors in a way that maintains a strategic distance from issues related with exceptionally connected factors. What's more, the correct use of

scales can advance the examination procedure by permitting the estimation and analysis of ideas that require more than single thing measures.

Q3) What guidelines can you use to determine the number of factors to extract? Explain each briefly.

Solution 3) The appropriate rules used depend to some degree upon the research question and what is thought about the quantity of factors that ought to be available in the information. If the researcher knows the quantity of factors that ought to be available, at that point the number to concentrate might be indicated in the start of the analysis by the from the earlier foundation. If the research question is generally to clarify a base measure of variance then the level of variance foundation might be generally imperative. At the point when the goal of the research is to decide the quantity of dormant factors hidden an arrangement of factors a blend of basis, conceivably including the from the earlier and level of variance measure, might be utilized as a part of choosing the last number of factors. The inactive root rule is the most generally utilized technique. This technique is to remove the quantity of factors having eigenvalues more prominent than

1. The rationale being that a factor ought to clarify in any event as much variance as a solitary variable. A related technique is the scree test model. To build up this test the idle roots (eigenvalues) are plotted against the quantity of factors in their request of extraction. The subsequent plot demonstrates an elbow in the inclined line where the remarkable variance starts to command basic variance. The scree test model for the most part shows a bigger number of factors than the dormant root run the show. One of these four model for the underlying number of factors to be separated ought to be indicated. At that point an underlying arrangement and a few trial arrangements are computed. These arrangements are turned and the factor structure is analyzed for importance. The factor structure that best speaks to the information and clarifies a worthy measure of variance is held as the last arrangement.

Q4) How do you use the factor-loading matrix to interpret the meaning of factors?

Solution 4) The initial step is to comprehend the factor-loading matrix and afterward distinguish the biggest huge loading of every factor as for a factor/set of factors. Further, moving evenly over the factor matrix and underlining the most elevated noteworthy loading for every factor. Once improved the situation each factor the researcher keeps on searching for other critical loadings. On the off chance that there is basic structure, just single critical loadings for every factor, at that point the factors are named. Factors with high factor loadings are viewed as more vital than factors with bring down factor loadings in the understanding stage. When all is said in done, factor names will be allocated so as to express the factors which load most altogether on the factor.

Q5) How and when should you use factor scores in conjunction with other multivariate statistical techniques?

Solution 5)

Analyst is keen on making a totally new arrangement of fewer composite variables to supplant either to a limited extent or totally the original arrangement of variables, at that point the analyst would register factor scores for use all things considered composite variables. Factor scores are composite measures for each factor speaking to each subject. The original raw data measurements and the factor analysis comes about are used to register factor scores for every person. Factor scores may reproduce as effortlessly as a summated scale, along these lines this must be considered in their utilization.

Q6) What are the differences between factor scores and summated scales? When is each most appropriate ?

Solution 6) The major contrast between the two is that the factor score is processed in view of the factor loadings of all variables loading on a factor, though the summated scale is figured by consolidating just chosen variables. In this manner, the factor score is described by not just the variables that load exceedingly on a factor, yet in addition those that have bring down loadings. The summated scale speaks to just those variables that load exceptionally on the factor.

Albeit both summated scales and factor scores are composite measures there are contrasts that prompt certain points of interest and hindrances for every strategy. Factor scores have the benefit of speaking to a composite of all variables loading on a factor. This is additionally a detriment in that it makes understanding and replication more troublesome. Likewise, factor scores can hold orthogonality while summated scales may not stay orthogonal. The key favorable position of summated scales is, that by including just those variables that load profoundly on a factor, the utilization of summated scales makes translation and replication less demanding. In this way, the choice decide would be that if data are utilized just in the original example or orthogonality must be kept up, factor scores are appropriate. In the event that generalizability or transferability is wanted at that point summated scales are favored.

Q7) What is the difference between Q-type factor analysis and cluster analysis

Solution 7) Both Q-Type factor analysis and cluster analysis contrast a progression of reactions with various variables and place the respondents into a few gatherings. The distinction is that the subsequent gatherings for a Q-type factor analysis would be founded on the intercorrelations between the methods and standard deviations of the respondents. In a commonplace cluster analysis approach, groupings would be founded on a separation measure between the respondents' scores on the variables being analyzed.

Q8) When would the researcher use an oblique rotation instead of an orthogonal rotation? What are the basic differences between them?

Solution 8) In an orthogonal factor rotation, the correlation between the factor axes is subjectively set at zero and the factors are thought to be autonomous. This streamlines the

mathematical methodology. In oblique factor rotation, the angles between axes are permitted to look for their own particular values, which rely upon the thickness of variable clusterings. In this way, oblique rotation is more adaptable and more reasonable (it takes into account correlation of fundamental measurements) than orthogonal rotation despite the fact that it is additionally requesting mathematically. Indeed, there is yet no accord on a best technique for oblique rotation.

At the point when the goal is to use the factor brings about a subsequent statistical analysis, the analyst may wish to choose an orthogonal rotation methodology. This is on the grounds that the factors are orthogonal (autonomous) and along these lines dispense with collinearity. In any case, if the analyst is basically keen on getting hypothetically important develops or measurements, the oblique factor rotation might be more attractive on the grounds that it is hypothetically and exactly more practical.

References :

Textbook : Multivariate Data Analysis

Book by Barry J. Babin, Hair, Rolph E Anderson, and William C. Black