ISBN: 978-81-971650-3-0

12. Multivariate Analysis Techniques

Dr. T. M. Prashanth Kumar

Assistant Professor,
Department of Studies in Education,
Vijayanagara Sri Krishna Devaraya University,
Ballari, Karnataka.

Abstract:

Multivariate analysis (MVA) approaches allow more than two variables to be evaluated simultaneously. There are two general types of MVA techniques: dependency analysis and interdependence analysis. Techniques are chosen based on the type of data and the reason for the study. Cluster analysis: "Techniques for identifying separate groups of similar cases". Data can also be summarized by dividing it into segments of comparable cases. Multivariate analysis includes a variety of techniques for analyzing data. In this study, we will discuss various strategies in terms of their usefulness and difficulty. And we provide an executive understanding of these multivariate analysis techniques, resulting in an understanding of the appropriate applications for each, assisting researchers in understanding the types of research questions that can be formulated, as well as the capabilities and limitations of each technique in answering those questions. In this paper, we will discuss. Multivariate analysis techniques.

Keywords:

Multivariate, Analysis, Techniques, Variables, Dependence, Interdependence., Cluster Analysis, Research Questions, Reduce Data, Construct, Test Hypotheses, Multiple Linear Regression, Multiple Logistic Regression

12.1 Introduction:

Multivariate analysis refers to a set of approaches that can be applied when several measurements are taken on each human or object in one or more samples. We will refer to the measurements as variables, and the individuals or objects as units (research, sample, or experimental) or observations. In practice, multivariate data sets are widespread, although they are not usually evaluated as such. However, the exclusive use of univariate processes with such data is no longer acceptable, given the availability of multivariate approaches and low-cost computing capacity to perform them. [1]

Historical Perspective: In 1889, Galton introduced the normal distribution, which was employed in traditional statistical methods to produce the correlation coefficient and linear regression. Fisher proposed analysis of variance and discriminant analysis in the 1930s, SS Wilkes developed multivariate analysis of variance, and H. Hotelling established principal component analysis and canonical correlation. In general, the majority of multivariate analysis theory was developed during the first part of the twentieth century.

60 years later, with the advancement of computer science, psychology, and multivariate analysis tools, the study of many other fields has become more widespread. SAS and SPSS, which were previously only available on mainframes, are now available for Windows. The marketing research analyst can now study data using a far broader range of sophisticated methodologies. The problem is deciding which technique to use and fully understanding its strengths and disadvantages. [2]

Initially, multivariate test and analysis methods were employed in statistics to identify causal correlations. Because human calculations are extremely complex, the approaches were only made practical in other domains of application after the introduction of related gear and software. Multivariate analysis methods are utilized in a variety of fields:

- Linguistics, Natural Sciences, and Humanities.
- Economy, insurance, and financial services.
- Data mining, large data, and relational databases.

Multivariate studies are typically performed using software to cope with large volumes of data and monitor the changed variables in practical applications such as usability tests. However, multivariate tests can also help to increase user friendliness on a smaller scale.

Multivariate analysis is a technique for analyzing several variables at the same time. Its purpose is to identify patterns, correlations, and associations among variables. In contrast to univariate analysis, which focuses on a single variable, multivariate analysis investigates the interplay of several factors. [3]

Multivariate analysis is fundamentally a quantitative approach to decision-making. It has a wide range of applications, including engineering, traffic management, biology, economics, marketing, ethics, and behavioural psychology. You can quantify how changes in one or more areas of a complicated problem affect a result over time and determine whether those adjustments will help or worsen the situation.

Typically, this form of analysis aims to accomplish the following goals:

- **Reduce data.** The technique allows researchers to reduce large amounts of data into more readable representations.
- **Simplify structure.** The analysis simplifies the structure of complex data sets, making them easier to interpret and use.
- **Sort of group trends and data.** This analysis is used by researchers to organize data into groups or trends so that it may be used for its intended purpose more easily.
- **Identify dependencies among variables.** Researchers use multivariate data to uncover particular dependencies between data sets in order to better comprehend the interactions between them.
- **Predict relationships between variables.** This study aids in predicting future linkages between data sets, as well as the creation of new data as variables change.
- **Construct and test hypotheses.** This analysis enables researchers to develop and test hypotheses about the linkages between data sets, trends, and potential data in order to advance their research. [4]

12.2 An Example of Multivariate Analysis:

Assume you're interested in the association between a person's social media use and their self-esteem. You might conduct a bivariate analysis by comparing the following two variables.

How many hours a day does a person spend on Instagram?

Their self-esteem score (determined by a self-esteem scale)

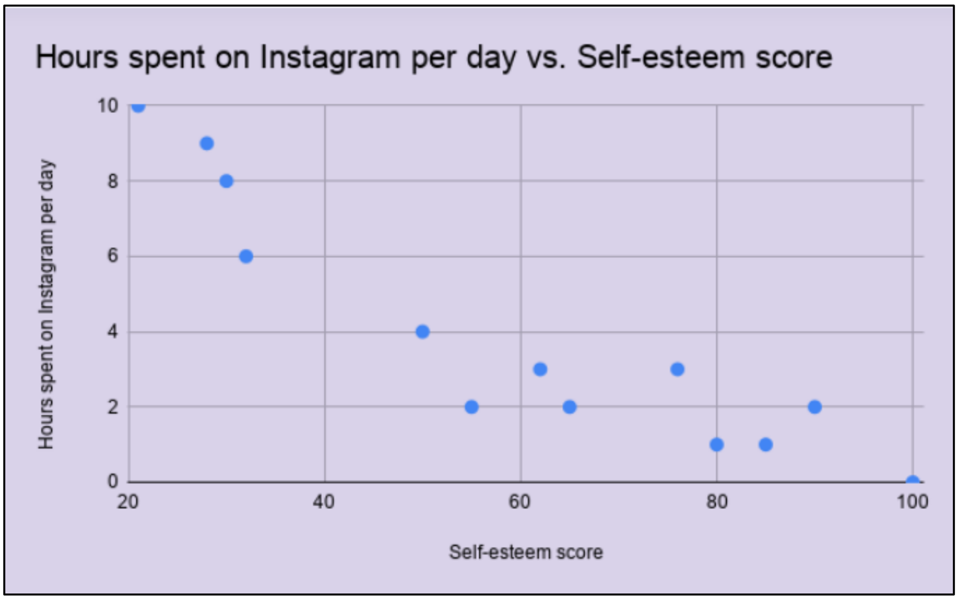

Figure 12.1: Example [5]

A. Level of Measurement and Multivariate Statistical Technique:

Independent Variable	Dependent Variable	Technique
Numerical	Numerical	Multiple Regression
Nominal or Numerical	Nominal	Logistic Regression
Nominal or Numerical	Numerical (censored)	Cox Regression
Nominal or Numerical	Numerical	ANOVA, MANOVA
Nominal or Numerical	Nominal (2 or more values)	Discriminant Analysis

B. Multivariate Methods can be Classified Based on The Types of Variables:

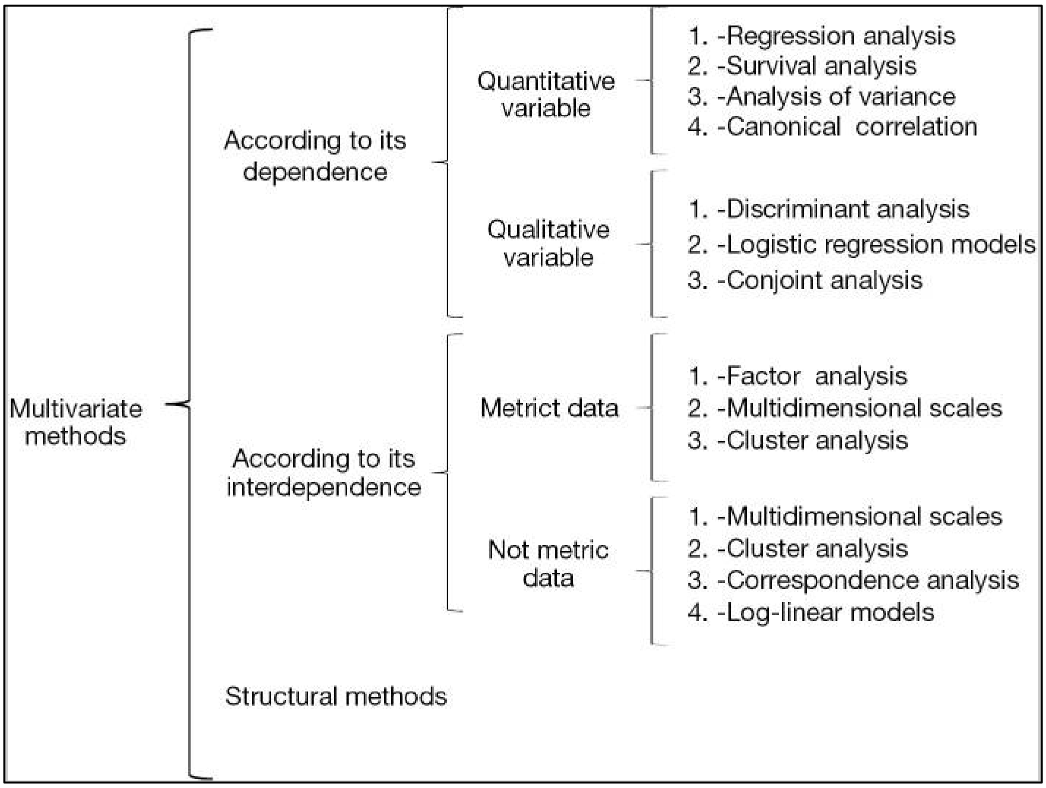

Figure 12.2: Classification of multivariate methods. [6]

C. Multivariate Analysis Techniques: Dependence Vs. Interdependence

Dependence Methods:

Dependence methods are employed when one or more variables rely on others. Dependence investigates cause and effect; that is, can the values of two or more independent variables be used to explain, characterize, or forecast the value of a dependent variable? For example, the dependent variable "weight" could be predicted by independent variables "height" and "age." [7]

Interdependence Methods:

Interdependence approaches are used to understand the structural makeup and underlying patterns of a dataset. In this situation, no variables are interdependent, thus you are not looking for causal interactions. Rather, interdependence techniques attempt to give meaning to a collection of variables or to group them together in meaningful ways.

With that in mind, let's look at some practical multivariate analysis techniques. We'll look into:

- Multiple linear regression
- Multiple logistic regression
- Multivariate analysis of variance (MANOVA)
- Factor analysis
- Cluster analysis [8]

Multiple Linear Regression:

Multiple linear regression is a dependent technique that examines the relationship between a single dependent variable and two or more independent variables. For example, suppose a couple wishes to sell their house. The price they can receive for it is determined by a number of independent variables, including location, time of year, interest rates, and the presence of comparable listings in the region.

Example of Multiple Regression:

As a data analyst, you might utilize multiple regression to forecast crop growth. In this case, crop growth is the dependent variable, and you want to observe how various factors influence it. Rainfall, temperature, amount of sunlight, and amount of fertilizer put to the soil are all possible independent variables. A multiple regression model will show you the percentage of variance in crop growth that each independent variable accounts for.

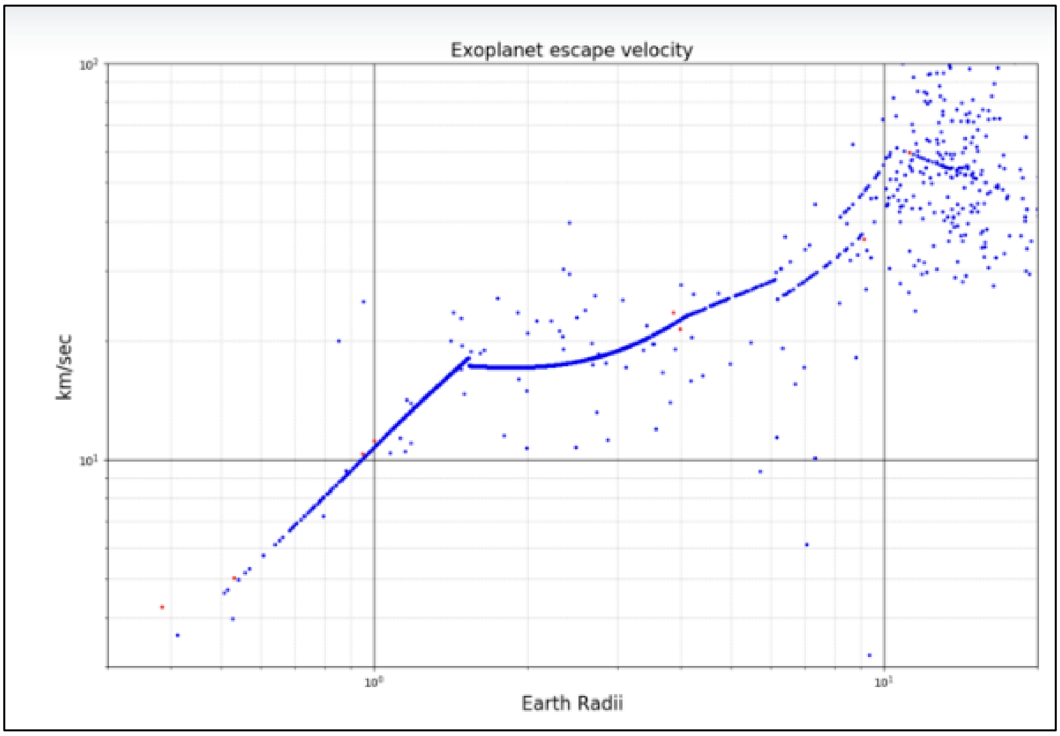

Figure 12.3: Example of multiple regression [9]

Multiple Logistic Regression:

In contrast, multiple logistic regressions look at the likelihood of one of two occurrences occurring. Will one's partner accept a marriage proposal? Will a youngster graduate from college? In both circumstances, a variety of factors may influence the final choice, including whether the partner had previously expressed interest in marriage. Does the child study hard? Finally, both boil down to a simple yes or no.

12.3 Multivariate Analysis of Variance (MANOVA):

This technique investigates the relationship between many categorical independent factors and two or more metric dependent variables. Whereas analysis of variance (ANOVA) compares differences across groups (using T tests for two means and F tests for three or more means), MANOVA investigates the dependence relationship between a collection of dependent measures across groups. This analysis is typically employed in experimental design, with a hypothesis on the relationship between dependent measures. This technique differs slightly in that the independent variables are categorical while the dependent variable is metric. Sample size is a concern, with 15-20 observations required each cell. However, if there are too many observations per cell (more than 30), the approach loses practical value. Cell sizes should be nearly comparable, with the largest cell having fewer than 1.5 times the observations as the smallest cell. That is because the dependent variables' normalcy is critical in this technique. The model fit is measured by comparing mean vector equivalents between groups. If there is a substantial difference in means, the null hypothesis can be rejected and treatment differences identified. [10]

Example of MANOVA:

Assume you work for an engineering firm that is on a mission to develop a super-fast, environmentally friendly rocket. You may use MANOVA to determine the influence of different design combinations on the rocket's speed and carbon dioxide emissions. In this instance, your categorical independent variables may be:

- Engine type, categorized as E1, E2, or E3
- Material used for the rocket exterior, categorized as M1, M2, or M3
- Type of fuel used to power the rocket, categorized as F1, F2, or F3

Your metric dependent variables are speed (km/h) and carbon dioxide (ppm). Using MANOVA, you would test various combinations (e.g., E1, M1, and F1 vs. E1, M2, and F1, vs. E1, M3, and F1, and so on) to determine the effect of all independent variables. This should help you determine the best design solution for your rocket. [11]

12.3.1 Factor Analysis:

When there are many variables in a research design, it is generally beneficial to condense the variables into a smaller collection of components. This is an independence technique, meaning there is no dependent variable. Rather, the researcher seeks to understand the fundamental structure of the data matrix.

Ideally, the independent variables are normal and continuous, with three to five variables loading onto each component. The sample size should be more than 50 observations, with at least five each variable. Multicollinearity is often desirable between variables, as correlations are critical for data reduction. Kaiser's Measure of Statistical Adequacy (MSA) assesses the extent to which one variable may be predicted by all other variables. An overall MSA of.80 or greater is considered extremely good, while a score of less than.50 is considered poor. [12]

12.3.2 Cluster Analysis:

The goal of cluster analysis is to reduce a big dataset to relevant subsets of people or objects. The divide is made on the basis of the items' similarity across a set of specified characteristics. Outliers are a concern with this technique, and they are frequently produced by an excessive number of irrelevant factors. The sample should be representative of the entire population, and uncorrelated characteristics are preferred. There are three types of clustering methods: hierarchical, which is a tree-like procedure ideal for smaller data sets; nonhierarchical, which requires a priori cluster number determination; and a combination of the two. There are four fundamental requirements for establishing clusters: the clusters must be unique, reachable, quantifiable, and profitable. This is an excellent resource for market segmentation.

Cluster analysis is a set of techniques used to categorize objects or instances into relative groups known as clusters. In cluster analysis, no prior information about any of the objects' group or cluster membership is provided.

- Cluster analysis involves partitioning data into groups based on similarity and assigning labels to each group.
- The primary advantage of clustering over categorization is its adaptability to change and ability to identify valuable traits that separate distinct clusters. [13]

12.3.3 Correspondence Analysis:

This technique allows for the dimensional reduction of object ratings on a set of attributes, yielding a perceptual map of the ratings. However, unlike MDS, both independent and dependent variables are analyzed simultaneously. This technique is closer in nature to factor analysis. It is a compositional approach that is beneficial when dealing with a large number of qualities and enterprises. It is most commonly used to analyze the efficacy of advertising efforts. It is also employed when the characteristics are too similar for factor analysis to be useful. The primary structural method is the creation of a contingency (crosstab) table. This suggests that the variables should be nonmetric. The model can be evaluated by calculating the Chi-square value. Correspondence analysis is challenging to comprehend since the dimensions include both independent and dependent variables.

12.3.4 Conjoint Analysis:

Conjoint analysis, often known as "trade-off analysis," allows for the evaluation of objects as well as the examination of many levels of qualities. It is both a compositional and a dependency approach, as it develops a preference for a set of traits and levels. A part-worth, or utility, is determined for each level of each characteristic, and combinations of attributes at certain levels are added to determine the overall preference for the attribute at each level. Models can be created to determine the optimal amounts and combinations of qualities for products and services.

12.4 Types of Variables:

A. Qualitative Variable:

- The measurement cannot be numerical.
- Observations are made when individuals are allocated to mutually exclusive categories.
- Non-numerical data.
- Observations, such as hair color and pathogen susceptibility, cannot be properly arranged or measured.

B. Quantitative Variable:

- Observations can be measured.
- Observations have a natural ranking.
- Observations have numerical values, such as yield, height, enzyme activity, etc.

Quantitative variables can be split into two categories:

a. Continuous: A variable that can take any value within a given range.
b. Discrete: One in which all values in a range are not conceivable, commonly used for counting data (number of insects, lesions, etc.). [14]

12.5 Conclusion:

Each multivariate technique outlined above is best suited to a specific sort of research inquiry. Each technique has its own set of strengths and shortcomings, which the analyst should be aware of before attempting to interpret the data. Any of the above procedures has strengths and shortcomings, which means that the analyst must exercise caution when employing these techniques, understanding the strengths and weaknesses of each. Each multivariate technique describes a different type of data than the other techniques. Statistical applications such as SPSS, SAS, and others make it simple to perform a technique.

12.6 References:

1. Aiken, L. S., Stein, J. A., & Bentler, P. M. (1994). Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. Journal of Consulting and Clinical Psychology, 62(3), 488.
2. Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster analysis: Quantitative applications in the social sciences. Beverly Hills: Sage Publication.

3. Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). Applied multiple regression/correlation analysis in the behavioral sciences (Third Edition). Mahwah, NJ: Erlbaum.
4. Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. Psychological Bulletin, 114(1), 174.
5. Ding, C., & He, X. (2004, July). K-means clustering via principal component analysis. In Proceedings of the twenty-first international conference on Machine learning (p. 29). ACM.Huberty, C. J. (1994). Why multivariable analyses? Educational and Psychological Measurement, 54, 620- 627.
6. Dunteman, G. H. (1989). Principal component analysis. Quantitative applications in the social sciences series (vol. 69). Thousand Oaks, CA: Sage.
7. Izenman, Alan J. (2008). Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer Texts in Statistics. New York: SpringerVerlag. ISBN 9780387781884.
8. Jump up to:a b Olkin, I.; Sampson, A. R. (2001-01-01), "Multivariate Analysis: Overview", in Smelser, Neil J.; Baltes, Paul B. (eds.), International Encyclopedia of the Social & Behavioral Sciences, Pergamon, pp. 10240–10247, ISBN 9780080430768, retrieved 2019-09- 02
9. "Handbook of Applied Multivariate Statistics and Mathematical Modeling | ScienceDirect". Retrieved 2019-09-03.
10. Schafer, J. L. (1997) Analysis of Incomplete Multivariate Data. CRC Press. (Advanced)
11. Sharma, S. (1996) Applied Multivariate Techniques. Wiley. (Informal, applied)
12. Grimm LG, Yarnold PR. eds. Reading and understanding multivariate statistics. Washington, DC: American Psychological Association Washington, 2011.
13. Alvin C. Rencher.2002: " Methods of Multivariate Analysis ", A. John Wiley & sons, inc. publication, Second Edition.
14. Bryant and Yarnold. 1994: "Principal components analysis and exploratory and onfirmatory factor analysis". American Psychological Association Books. ISBN 978-1-55798-273-5.