Business Research Methods: Theoretical Demystification of The Use of Multivariate Analysis Techniques in Research

Austin Mwange, PhD/DBA
Senior Lecturer, Department of Economics, School of Social Sciences, ZCAS University, Lusaka, Zambia; Lecturer Graduate School of Business, University of Zambia, Lusaka, Zambia; Lecturer, University of Lusaka, Lusaka, Zambia
Email: austin.mwange@unza.zm; austin.mwange@zcasu.edu.zm; austin.mwange@unilus.ac.zm

Joseph Chiseyeng'i, Ph.D.
Doctorate in Business Administration Student, Department of Business Administration, School of Business, ZCAS University, Lusaka, Zambia; Senior Lecturer, Business Department & Zambia University College of Technology (ZUCT), Ndola Email Address: bashikapembe@gmail.com

Windu Matoka, DBA
Senior Lecturer, Department of Business Administration, School of Social Sciences, ZCAS University, Lusaka, Zambia. Email: windumatoka@gmail.com; windu.matoka@zcasu.edu.zm

Abstract
Multivariate data analysis techniques observe and analyse multiple statistical variables. These are more advanced than univariate methods (methods that analyse one variable) and bivariate methods (methods that analyse two variables). Multivariate analysis methods were developed to analyse datasets containing multiple variables simultaneously and are ideal for analysing large datasets to reveal causal and effect relationships between variables. This paper identifies different categories of multivariate analysis methods, discusses the assumptions they are based on, analyses their goals and objectives, and describes their advantages and limitations. It also discusses the factors researchers must consider when determining the best technique for a particular research project. The article concludes with a discussion of commonly used methods of multivariate data analysis. This is not a discussion of the statistics underlying each method, but rather an introduction to multivariate methods and their capabilities and limitations in answering research questions.
Keywords: Multivariate analysis, Univariate analysis, Bivariate analysis, ANOVA, MANOVA,
DOI: 10.7176/JEP/14-21-06
Publication date: July 31st 2023

1. Introduction
The main purpose of research is for researchers to collect and analyse data and draw research conclusions. Data analysis is the process of systematically applying logical and statistical techniques to collect, describe, summarise, analyse, and evaluate data to facilitate informed decision making. Researchers can draw inductive inferences from population data using a variety of statistical and analytical techniques (Shamoo and Resnik, 2003). The data analysis process is often a continuous, iterative process in which data is continuously collected and analyzed simultaneously. The form of data analysis is determined by the type of data collection, research methodology, and research question (Savenye and Robinson, 2004).

There are three levels of data analysis. First, at the univariate level, the researcher analyses only one of his values. Second, it occurs at the bivariate level, where the researcher analyses only her two variables. Finally, data analysis is done at the multivariate level, where multiple variables are analysed simultaneously in one study. This discussion focuses on goals, objectives, assumptions, and limitations of multivariate data analysis.

2. Methodology
The methodology used in this study was a combination of theoretical and narrative literature review. The research focus was on multivariate techniques. The articles eligible for this study were drawn primarily from statistics textbooks and academic papers from Google Scholar, EBSCO, and JSTOR.

3. Discussion
3.1. Defining Multivariate Data Analysis
There are three levels of data analysis in statistics:
- **(a)** *Univariate analysis* -analyses just one variable
- **(b)** *Bivariate analysis* - analyses two variables
- **(c)** *Multivariate analysis* – analyses more than two variables

Whether researchers use univariate, bivariate, or multivariate methods depends on the type of data, number of variables, research focus, and research question.

The univariate method is the simplest method of data analysis. As the name suggests, Uni means 1, and in univariate analysis there is only one reliable variable that researchers examine (Clark, 2011). Univariate analysis methods are used by researchers when the data contains only one variable and the analysis does not attempt to reveal causal relationships. It is the most basic form of statistical data analysis, whose purpose is to describe and summarise data and discover underlying patterns. This is often done by computing statistics such as mean, median, mode, spread, variance, range and standard deviation. An example of univariate data analysis is a researcher measuring the height of students in a particular class. In this experiment, the data simply reflect one variable, weight and its quantity. Univariate data analysis is often represented in various forms such as frequency tables, polygons, histograms, bar charts, and pie charts.

The next level of data analysis after univariate analysis is bivariate analysis. In bivariate data analysis, a researcher examines two variables and the analysis relates to causal relationships between variables (Steven, 2001). Bivariate bivariate data analysis is more analytical than univariate univariate data analysis. An example of bivariate data analysis is when a researcher obtains the weight and height of all students in a particular class and further investigates whether there is a relationship between the weight and height of the students. We then apply bivariate data analysis methods to determine the correlation between weight and height. Bivariate analysis is performed using correlation coefficients and regression analysis.

So-called multivariate data analysis techniques follow univariate and bivariate data analysis methods. Multivariate analysis allows researchers to measure multiple variables in complex experimental settings and determine the impact of each variable on others. They help researchers quantify and understand relationships between variables in a dataset (Johnson, 2002). Multivariate analysis is a more complex form of statistical analysis and a more advanced statistical technique than univariate and bivariate methods. Roy, Fischer and others pioneered multivariate analysis and laid the foundation for the statistical analysis methods that are so common today (Hair, 2010). This method is a very useful tool for researchers who want to analyse and interpret large amounts of data. Multivariate analysis is a highly complex and time-consuming data analysis method that, while useful, requires researchers to be proficient in its application and understanding.

Multivariate data analysis reflects what is actually available. An example of its application is when a researcher collects data on a student's weight, height, cholesterol levels, and blood pressure. We will also collect data on student diet and examine the relationship between the four variables and student diet. This is an example of a study that requires the use of multivariate data analysis techniques. Common multivariate analysis techniques include multiple regression, cluster analysis, principal component analysis, multivariate analysis of variance (MANOVA), multidimensional scaling, factor analysis, and discriminant analysis. The following table shows the different types of multivariate analysis methods.

TABLE 1: Classification of Multivariate Data Analysis Methods

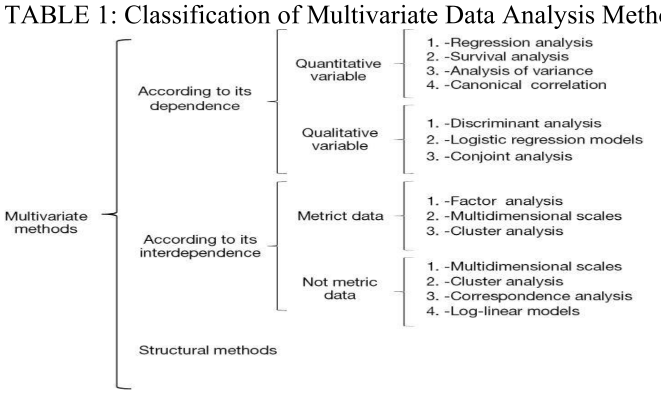

Table taken from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4387392/figure/f1/

3.2. Objective of Multivariate Analysis

Multivariate analysis is a type of exploratory data analysis that allows you to gain a deeper understanding of your data. It aims to identify patterns and correlations between multiple variables, enabling a deeper and more complex understanding of research questions that are difficult to obtain with univariate and bivariate analyses. Multivariate data analysis has several goals. A variable is a weighted combination of variables, and the main goal and purpose of multivariate analysis is to find the optimal combination of weights (Weisberg, 1985).

The first goal of multivariate methods is to reduce or structurally simplify data without sacrificing valuable information, thereby greatly simplifying data analysis and interpretation. A second goal is to facilitate sorting and grouping of data. When multiple variables are present, researchers should group similar variables to make data analysis more manageable. A third purpose is that multivariate analysis helps researchers identify dependencies between different variables. This answers the research question. Are one or more variables dependent on others, or are all variables independent of each other? A fourth goal of multivariate analysis is to allow researchers to predict the values of variables based on the revealed relationships between them. Finally, multivariate analysis aims to help researchers formulate and test theories and hypotheses (Hair, 2010).

3.3. Types of Multivariate Tests

Most of the problems we see in the real world are multivariate. That is, it combines multiple dependent variables to produce results. As an example, when a researcher wants to analyse and forecast the sales of a business unit, he has to consider various factors and variables. These factors include product quality, packaging, geographic location of sales locations, marketing efforts, market competition, production costs, and other variables. If researchers want to analyse how these factors affect sales volume, they need to apply multivariate analysis.

Multivariate analysis is independent of any particular method, but covers a wide range of statistical methods (Clark 2011). Researchers have access to over 20 different multivariate analysis methods, and choosing the most appropriate method depends on the nature of the data and the question of investigation.

Classification of Multivariate Techniques

Multivariate methods can be classified as either full or fractional, dependence or interdependence.

(a) *Full or Fractional* - A full factor, also called multivariate test, is a method of designing and testing all possible combinations of variables equally. If a researcher should test 3 variants for one element and 2 variants for another element, for each of the 6 combinations he would be assigned 17%. Increase. Fractional factors, on the other hand, use only a subset of all possible combinations.

(b) *Dependence or Interdependence* - Various multivariate analysis techniques can be divided into two categories: Dependency Techniques and Interdependence Techniques. The use of the terms "dependency" and "interdependency" refer to the different types of relationships contained in a data set.

Dependency methods are multivariate methods used when one or more variables are determined to be dependent on other variables. Dependent methods examine cause and effect and attempt to determine whether the values of two or more independent variables can describe, explain, or predict the values of another dependent variable. Interdependency methods are used to group variables in meaningful ways to reveal underlying patterns and structural organisation of data sets. Researchers do not look for causality because no variable depends on other variables (Clark, 2011). Therefore, multivariate dependence methods are about the influence of certain variables on other variables, and multivariate interdependence methods are about the structure of the dataset.

3.4. Advantages and Limitations of Multivariate Analysis

Multivariate analysis analyses data that occurs in the real world. One of the major advantages of studying multiple variables is the depth of insight provided to the researcher. In practice, most events cannot be traced back to a single factor, so multivariate methods attempt to model reality (Yarnold 2011). For example, a smartphone purchase decision can include the variables price, brand name, colour, and features.

The advantage of multivariate methods is that they allow researchers to analyse different factors and variables, making research conclusions more realistic and closer to real situations. The conclusions researchers can draw are more accurate because multiple factors of the independent variables that affect the variability of the dependent variable are taken into account. Therefore, multivariate methods provide meaningful tests of significance compared to univariate or bivariate methods. After the initial benefits, multivariate methods can help researchers save a lot of time and resources. Multivariate methods allow researchers to analyse multiple tests simultaneously, eliminating the need to perform multiple univariate or bivariate analyses.

Another advantage of multivariate methods is that they allow researchers to identify and quantify associations between variables over a wide range. You can also manipulate and control associations between variables using cross tabulation, partial correlation, and multiple regression, introduce other variables into the model, and determine relationships between independent and dependent variables. . Finally, another advantage of multivariate methods is that they are of particular importance in the social sciences, where randomised laboratory experiments are not applicable. In these experimental cases, multivariate methods allow statistical estimation of relationships between different variables. They can relate the importance of each to revenue and identify dependencies that exist between them (Hair, 2010).

Before they appear to be a data analysis panacea, it is important to highlight the limitations of multivariate data analysis that hinder their application in specific cases. Researchers should be aware of these limitations, especially when deciding which method to use in a particular research scenario. The main drawback of multivariate methods is that researchers must test large amounts of data in order to obtain meaningful results. When testing a significant number of variables, researchers must contend with a huge number of possible combinations. In practice, this leads to unnecessarily long and complex tests and means that the necessary statistical confidence in decision-making cannot be achieved (Christennsen, 2012).

A second drawback of multivariate methods is that they require rather complex statistical and mathematical calculations to obtain satisfactory research results. Running univariate or bivariate tests that require only one or two variables is a fairly simple case. Researchers do not need to perform complex statistical manipulations to try to understand the relationship between two variables. The problem arises when researchers study relationships between multiple variables. Multivariate methods are very complex and often require sophisticated and expensive statistical programs to analyse the data (Pascual, 2003). Moreover, multivariate methods and their statistical models are not always easy for researchers to understand and interpret. They are so complex and difficult to understand and apply that there is always a high risk that studies will misapply multivariate methods.

Finally, another drawback of multivariate methods is that they are time consuming, as a large number of variables need to be collected, summarised, and analysed. A large sample size is required for multivariate methods to produce meaningful results and reduce standard errors (Hair, 2010).

3.5. Choice of Multivariate Techniques

Choosing an appropriate multivariate technique depends on several factors, including whether the variables are independent or dependent. Researchers use dependency methods when related variables are dependent on each other. Dependency method is a multivariate analysis technique used when one or more variables are identified as dependent and the remaining variables are independent. On the other hand, researchers use the interdependence method when the variables are independent of each other. If variables are dependent, you need to determine how many variables are dependent in one analysis. Once this has been determined, researchers need to know how to measure both the dependent and independent variables (Hair, 2010).

In addition to dependence and independence, there are other factors that influence the choice of multivariate method that researchers can consider. First, each multivariate method should be evaluated in terms of the type of research question for which each multivariate method is best suited. Each multivariate method has a specific type of research question that it is best suited for. For example, multiple regression is used when researchers try to establish relationships between variables. This multivariate technique cannot be applied to research questions in which the variables are independent and unrelated. Similarly, researchers should consider the intent and goals of any multivariate method used. Another factor to consider is the structure of your data. This strongly influences the selection of the appropriate multivariate method to use. For example, multiple regression cannot be applied to qualitative data. Researchers should also consider sampling techniques suitable for a particular multivariate method. Qualitative data requires qualitative data sampling methods such as convenience, snowballing, and other sampling methods. Similarly, the applicable sample size also influences the type of multivariate method used. Finally, researchers should specify a mathematical or statistical model for each multivariate method (Clark, 2011).

Steps in Multivariate Data Analysis

Mengual-Macenlle (2015) outlines the general steps that the researcher must consider when planning to perform a multivariate analysis. The researcher must:

- **(a)** First define the objectives of the research analysis. The research problem must be defined in its conceptual terms, objectives and multivariate techniques that are going to be applied
- **(b)** Then design the research analysis. S(he) must determine the sample size and consideration of the estimation techniques to be applied
- **(c)** Decide what to do with missing data if any
- **(d)** Now perform the data analysis. S(he) needs to identify influential observations and any outliers that might have impact on the data analysis process
- **(e)** Interpreting the results of the analysis, which might require him or her to redefine the model and the variables; that is revert back to steps (iii) and (iv) above
- **(f)** Validate the research results.

Mengual-Macenlle (2015) goes on to provide hints on how to apply Multivariate Analysis. He advises researchers that when the results of the analysis don't seem to make sense, then the researcher must consider the reliability of the analysis. He then provides the following guidelines to the researcher:

- **(a)** To be as specific as possible in what the researcher want to analyse and interpret
- **(b)** Base the analysis on actions the researcher can practically take or decisions s(he) can practically make
- **(c)** Remove from the analysis variables the researcher can't manipulate or control
- **(d)** Must ensure that the data the researcher uses is reliable. The more important the decision to be made, the more reliable the data must be
- **(e)** The researcher must not read into the analysis than the analysis and report reveals. If the analysis produces a large margin of error, the researcher must take this into consideration

As the researcher goes through the process of identifying the appropriate multivariate method to apply, s(he) must consider the potential for complementary use of the methods. Multivariate data analysis is often a combination of techniques that together helps the researcher analyse the data. For instance, factors analysis can follow after cluster analysis, and finally the data can be subjected to a multiple regression analysis. The choice of the most appropriate method must be done with full realisation that each method has its strengths as well as its weaknesses.

4. Multivariate Data Analysis Methods

We now turn our focus to discussing some of the useful multivariate analysis methods that are available to the researcher.

4.1. Multiple Regression

Multiple Regression is by far the most widely used and versatile dependence technique applicable to every research and decision making. It is a general statistical technique used to analyse the relationship between a single variable and several independent variable. While simple regression maps one variable as a function of the other, multiple regression maps one dependent variable as a function of several other independent variables. Multiple regression shows the researcher the extent to which each independent variable has a relationship with the dependent variable. It is the foundation of most forecasting models that is often used to help the researcher understand which factors are most likely going to influence a certain outcome.

The purpose of regression analysis is to predict a single dependent variable from knowledge of one or more independent variables. Predict a single dependent variable using multiple independent variables with known values. To use it, researchers need to divide the variables into dependent and independent variables, and the variables must be quantitative and econometric.

Multiple regression analysis is based on the assumption that there is a relationship between the dependent and independent variables. Calculating the regression coefficients informs the researcher whether this assumption is met (Hair, 2010). A practical example of multiple regression is the study of the factors that determine tomato growth. Researchers may wish to examine how soil type, rainfall, temperature, insolation, and fertiliser levels affect tomato plants. Shows researchers the percentage of variance in plant growth.

4.2. Multiple Logistic Regression

When choosing an appropriate analytical method, researchers may encounter problems involving one qualitative dependent variable and several qualitative independent variables. In this case, researchers should use multiple logistic regression methods. Multiple logistic regression, also called choice model, is a type of multiple regression that allows prediction of binary events. It is used to determine the probability of binary events with only two possible outcomes (Clark, 2011). Binary events either fire the event or they don't. Based on a combination of independent variables, multiple logistic regression estimates the likelihood of a particular event occurring.

An example of multiple logistic regression is when a researcher analyses the likelihood that a potential customer will purchase a particular smartphone. Researchers may consider many independent variables, such as income level, age, gender, occupation, and location. Using these independent variables as inputs to a multiple regression model, researchers can calculate the probability that a potential customer will use a smartphone.

4.3. Multivariate Analysis of Variance (MANOVA)

At the bivariate level, analysis of variance (ANOVA) allows researchers to analyse the means of multiple groups, not just two groups. Multivariate analysis of variance (MANOVA), on the other hand, is used to measure the influence of multiple independent variables on two or more dependent variables. In this method, researchers analyse different combinations of independent variables to compare how they differ in their effect on the dependent variable (Hair, 2010). It is important to note that in MANOVA the dependent variable is quantitative and scale in nature, whereas the independent variable is qualitative and categorical.

An example of MANOVA is studying how different factors, such as rainfall, sun exposure, soil type and amount of fertiliser, affect both the growth of tomatoes and the number of fruits they produce. There are some researchers who In this example, the qualitative and categorical independent variables are:

- **(a)** Rainfall – categorised as Rain-1, Rain-2, Rain-3
- **(b)** Type of soil - categorised as Soil-1, Soil-2, or Soil-3
- **(c)** Amount of fertiliser – categorised as F1, F2 and F3

Quantitative and metric dependent variables are plant growth in centimetres and its productivity measured in the number of fruits per plant. This helps researchers find the best combination of soil type, rainfall, solar radiation and amount of fertiliser.

An advantage of MANOVA is that it is useful in research and experimental situations where at least some of the independent variables can be manipulated or controlled. The MANOVA method helps prevent Type I hypothesis errors that can occur when multiple ANOVAs are run independently. However, compared to bind ANOVA, although time-saving, MANOVA is an advanced and complex method. In addition, MANOVA uses several discriminant functions that are difficult for researchers to understand and interpret (Clark, 2011).

4.4. Factor Analysis

Factor analysis is an interdependent method aimed at reducing the number of variables in a dataset. Too many variables make it difficult to find patterns in your data. Researchers try to reduce the number of variables to reduce overfitting. Overfitting is a modelling error that occurs when a model fits too closely to one dataset and fails to fit another dataset, which can lead to poor prediction accuracy. Factor analysis works by condensing all highly correlated variables into a single variable. Therefore, it is common for researchers to perform factor analysis to prepare data for further analysis (Steven, 2001).

Suppose a researcher is analysing data about an individual's education level, occupation, and income level. He/she may perceive that these three factors are highly correlated and thus may combine them into one of her variables, such as the subject's socioeconomic status. Increase. Even if a researcher were to condense a set of variables into one of hers, no information would be lost.

Factoring a multivariate procedure is useful when the goal is to reduce the number of variables by combining one or more variables into a single variable. Another benefit is time savings. However, the usefulness and practicality of factorization methods depend on the researcher's ability to accurately identify and group related variables according to their properties. Moreover, this method can be highly misleading if invalid or unreliable data is provided (Hair, 2010).

4.5. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a method of computing the principal components of a dataset and using them to perform baseline changes in the data. This method effectively reduces overfitting, helps remove correlated features, and makes the data analysis process time efficient (Stevens, 20021). To apply PCA, the data to be analysed must be standardised. Otherwise, the independent variables will be difficult to interpret and the test will not be able to find the best principal components.

4.6. Cluster Analysis

Cluster analysis is another interdependence technique that researchers use to group similar items in a dataset into clusters. This analysis helps researchers find patterns and understand how the data is distributed. The purpose of cluster analysis is to include variables in one cluster that are more similar to each other than variables in other clusters. To achieve this, researchers need to measure both within-cluster and between-cluster distances. Within-cluster distance measures the distance between data points within a cluster. The distance within a cluster should ideally be small. Intercluster distance measures the distance between data points in different clusters, and this distance should ideally be large (Hair, 2010). Because cluster analysis is an interdependent method of data analysis, cluster analysis is often performed in the early stages of data analysis. An example of cluster analysis in action is market segmentation, which creates clusters of customers with similar characteristics.

When applying cluster analysis, the data collection process can be adapted to the method and type of research being conducted. This method can also be applied to analyse images, patterns, and ideas. The only drawback of cluster analysis is its inability to distinguish between related and irrelevant variables. Therefore, it is important for researchers to understand the relevant variables involved in clusters. Another drawback of cluster analysis is that it is time consuming and expensive (Stevens, 2001).

4.7. Discriminant Analysis

Discriminant analysis methods are used to classify and distinguish between two or more homogeneous data sets. They help researchers quickly analyse whether differences between data groups are significant. This method produces a discriminant function, known as the discriminant function, which is a linear function of the independent variables. Weights assigned to each independent variable are corrected for interrelationships between all variables. The weights are called discrimination coefficients. This method is best applied when the dependent variable is qualitative and the independent variable is quantitative with a high degree of normality (Hair, 2010).

Discriminant analysis allows researchers to classify ungrouped data based on the dependent variable. Its main weakness is that it is very sensitive to outliers and the dependent variable cannot be perfectly correlated with linear combinations of other variables (Clark, 2011).

4.8. Multidimensional Scaling (MDS)

Multidimensional Scaling (MDS) is a technique for creating a map showing the relative positions of multiple objects given a table of distances between them. This is known as the proximity matrix. Neighbourhood matrices or maps can consist of one or more dimensions and can arise directly from experiments or indirectly as correlation matrices. MDS programs compute quantitative or qualitative solutions. Generally, you should evaluate objects with at least four times the dimension. You can rate objects using qualitative preference ranking or qualitative similarity. Dimensions can be interpreted objectively by researchers or by having respondents identify dimensions (Johnson, 2002).

4.9. Correspondence Analysis

Discriminant analysis allows researchers to classify ungrouped data based on the dependent variable. Its main weakness is that it is very sensitive to outliers and the dependent variable cannot be perfectly correlated with linear combinations of other variables (Clark, 2011).

Similar to factor analysis, this method is used to visualise table rows and columns of non-negative data as points on a map. For cross tabulation, this method can be viewed as describing the association between table rows and columns as measured by the Pearson chi-square statistic. Correspondence analysis shares some similarities with principal component analysis (Kumar, 2016).

4.10. Conjoint Analysis

Conjoint analysis is a multivariate technique specifically designed to understand how respondents form their preferences for objects, products, services, or ideas. The flexibility of this analysis means that it can be applied to almost any area where decision-making is considered. Conjoint analysis focuses on researchers' ability to theorise about voting behaviour. Therefore, many of its findings should be viewed primarily as exploratory, as they are directly attributable to key assumptions made during the research study design and conduct.

Conjoint analysis is based on the premise that any set of objects or concepts can be evaluated as a bundle of attributes. After determining the contribution of each factor to the consumer's overall rating, the researcher then defined the item or concept with the best combination of characteristics and determined the relative contribution of each attribute at each level to the item's overall rating. Contributions can be shown and quotes can be driven using the customer's judgement. Predict across objects with different sets of traits and segregate groups of prospects who give different importance to traits to define high- and low-probability segments.

Conjoint analysis, also known as multi-attribute configuration modelling, discrete choice modelling, statutory preference research, or trade-off analysis, is part of a broad set of trade-off analysis tools used to systematically analyse decision making. It helps determine whether subjects like different attributes of a product. This method is often used in marketing to determine customer preference for certain features over others. It helps determine how buyers value the various attributes that make up a single product (Hair, 2010). For example, car manufacturers use conjoint analysis to understand the combinations of attributes such as price, shape, colour, speed, and features that buyers prefer. A major drawback of conjoint analysis is that full profile and tradeoff methods are impractical for more than 10 attributes, whereas many conjoint analysis require studies that need to include more than 20 attributes.

4.11. Canonical Correlation Analysis

Canonical correlation analysis is considered the most flexible of the multivariate analysis methods. It is a multivariate extension of correlation analysis that analyses linear relationships between two sets of variables and is often used for data reduction and interpretation. Researchers use this method to analyse the correlations that exist between one set of dependent variables and another set of independent variables. It helps to condense relationships into fewer statistics while preserving key characteristics of the relationships. When the dependent variables are related and the independent variables are related, finding the relationship is difficult without techniques like canonical correlation. The rationale behind canonical correlation is very similar to principal component analysis.

Unlike multiple analysis of variance (ANOVA), canonical correlation uses quantitative independent variables and can also use qualitative categorical variables. Because this technique is less restrictive than any multivariate analysis technique, the assumptions are loose and the results should be interpreted with caution.

4.12. Structural Equation Modelling

Structural Equation Modelling (SEM) is a flexible approach to examining how things are related to each other. There are three main characteristics of SEM as follows:

- **(a)** The estimation of multiple and interrelated dependence relationship
- **(b)** An ability to represent unobserved concepts in these relationships and correct for measurement error in the estimation process, and
- **(c)** A focus on explaining the covariance among the measured items

Structural equation modelling (SEM) is a multivariate statistical analysis technique aimed at studying structural relationships. It is used to determine the assumed causal relationship between a set of dependent and independent variables. It is a very broad and flexible data analysis method as it is not a single method but a combination of related methods.

5. Conclusions

In this white paper, we have described various multivariate analysis techniques and provided a better understanding of the appropriate use of each technique. You learned that multivariate analysis is used to analyse data containing more than one variable to reveal patterns and correlations between multiple variables. Multivariate analysis is especially useful for analysing complex data sets, allowing researchers to gain a deeper understanding of the data and how it relates to real-world situations. The two types of multivariate analysis techniques are: the dependency method, which examines causal relationships between variables, and the interdependence method, which examines the structure of data sets. Major multivariate analysis techniques include multiple linear regression, multiple logistic regression, MANOVA, factor analysis, and cluster analysis. Each multivariate method has a specific type of research question that it is best suited for.

6. References

Afifi A, May S, Clark VA. *Practical Multivariate Analysis*. Boca Raton, FL: CRC Press, 2011.

Clark (2012), *Practical Multivariate Analysis.* CRC Press, Taylor and Francis Group, Boca Raton, FL.

Gil Pascual JA. (2003), *Methods worth of research in education. Multivariate Analysis*

Grimm LG, Yarnold PR. eds. (2011), *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association Washington

Hair J.F. (2010), *Overview of Multivariate Methods.* Pearson Prentice Hall

Johnson R. A. et. al. (2002), *Applied Multivariate Statistical Analysis.* Upper Saddle River, NJ: Prentice Hall

Mengual-Macenlle, N., Marcos, P. J., Golpe, R., & González-Rivas, D. (2015). Multivariate analysis in thoracic research. *Journal of thoracic disease*, *7*(3), E2–E6. https://doi.org/10.3978/j.issn.2072-1439.2015.01.43

Rencher AC, Christensen WF. eds. (2012), *Methods of Multivariate Analysis.* New Jersey: Wiley.

Stevens J. (2001), *Applied Multivariate Statistics for Social Sciences;* Hillsdale, NJ: Lawrence Erlbaum Publishing

Weisberg S. (1985), *Applied Linear Regression;* New York. Wiley

Shamoo, A.E., Resnik, B.R. (2003). Responsible Conduct of Research. Oxford University Press.

Shepard, R.J. (2002). Ethics in exercise science research. Sports Med, 32 (3): 169-183.