Jurnal Pendidikan Matematika Vol: 3, No 1, 2025, Page: 1-16

An Overview of Multivariate Statistical Methods and Their Practical Applications

Abdulqader Mutlag Hamad

Department of Mathematics, college of Science, University of Mohaghegh Ardabili, Iran.

DOI:
https://doi.org/10.47134/ppm.v3i1.2084
*Correspondence: Abdulqader Mutlag Hamad
Email:
abdulqadirmatalkhamad@gmail.com

Received: 04-09-2025
Accepted: 18-10-2025
Published: 21-11-2025

Abstract: Multivariate data analysis is a powerful statistical approach used to analyze data involving multiple variables simultaneously. Researchers can use this method to find complicated ties, reduce the number of factors, and group data more effectively. When you need to understand data with more than one variable, you can use tools such as factor analysis, cluster analysis, discriminant analysis, principal component analysis, and multivariate regression. More and more fields, like business, engineering, health, and the social sciences, need multivariate analysis. This is because computers and other strong tools are getting better all the time. You will learn about some important multivariate methods and how they are used in the real world in this study. It also talks about the ideas that make them work. It talks about how these ways can help people make better decisions based on facts

Keywords: Multivariate Analysis, Factor Analysis, Cluster Analysis, Discriminant Analysis, Principal Component Analysis, Multivariate Regression, Dimensionality Reduction, Statistical Modeling, Mathematical Statistics

Introduction

We look at and figure out more than one result variable at the same time in a big part of math statistics called multivariate data analysis. One-variable techniques only let researchers look at one variable at a time, while multivariate techniques let them look at complex links between many factors at once. Their picture of how the data is set up is now more complete. In areas with a lot of interconnected factors that affect things, like engineering, health, the social sciences, and economics, this method works really well.

We study data that has a vector value in math. This means that each observation is made up of more than one measurement. In order to guess or decide what to do, the main goal is to describe and explain how these factors are linked on a deeper level. They should also look for underlying trends. Some of the tools that are used in this field are multivariate regression, factor analysis, principal component analysis, cluster analysis, discriminant analysis, and canonical correlation.

This is because as computers get faster and more specialized statistical software is made, multivariate methods become a lot more useful. This makes them faster and better able to work with big, hard data. These new tools let researchers carefully test their ideas

about how different things are connected. They can also organize data, cut down on the number of variables, and find secret trends that were hard to find with older tools that only used one variable.

When you look at multivariate data, you should also pay close attention to ideas like multivariate normality, regression, and homogeneity of differences to make sure you get the right results. If you use these methods correctly, they will help you figure out what data means. Scientists use them outside of the lab and they help them make decisions.

Historical Overview

Frans Galton came up with what most people do to break things up in 1889. This is how numbers are used now in many ways. The linear regression and the correlation coefficient are also his work. We use them every day because they are very useful statistics tools.

He came up with the ideas of discriminant analysis and analysis of variance (ANOVA) in the 1930s. These were very important to the field. Around the same time, S. S. Wilks made progress in the field by creating MANOVA, which stands for "multivariate analysis of variance." Harold Hotelling was the first person to use both PCA and CCA at the same time.

Most of the main ideas behind multivariate analysis were set by the middle of the 20th century. And as computer science grew over the next few decades, these stats tools became easier to use. They could also be used in more areas, like psychology and the social sciences.

Researchers can use SAS and SPSS to do a lot of different types of advanced studies. This is very true for people who do study on marketing. It used to be that you could only use these tools on mainframe computers. They can now also be used on normal computers that run Windows. These days, getting somewhere isn't the point; the point is to figure out which way to go and what its pros and cons are.

Introduction to Multivariate Statistical Analysis

Math and statistics are used in multivariate statistical analysis to make sense of problems with more than one variable or factor. Over the past two decades, the widespread use of computers and the growing need for data-driven solutions in both research and industry have led to the extensive application of these techniques in fields such as geology, meteorology, hydrology, medicine, industry, agriculture, and economics.

These methods provide effective solutions for real-world problems. Techniques like principal component analysis, factor analysis, and correspondence analysis are used to explore system structures by identifying key variables or subsets that capture the essence of a complex system. They help show the system as a whole and figure out how different things affect it.

Most tools for multivariate analysis can be put into two main groups:

1. Computer models that can guess A lot of the time, multiple linear regression, iterative regression, or a mix of these is used to try to guess what will happen.
   Descriptive models are one way to look through data for patterns or groups. Most of the time, they group things together using cluster analysis and other similar methods.

2. To group things that are alike and find new ones. In order to find links and trends in the organization, this is done. Tests used to be pretty simple and only had one variable. These tests didn't show how hard systems are in the real world. There are more advanced tools like discriminant analysis and cluster analysis that make it easier to sort numbers and make models these days.

Multivariate analysis: what it is and how it can be used

It's used to keep track of processes, learn about the market and users, make processes better, and do a lot more. It's not always easy to do controlled lab studies in the social sciences like it is in the hard sciences. These tips will really help in this case. Multivariate methods can help you find the mathematical links between these factors. They show how variables are connected and how much each one affects the outcome.

Why Use Multivariate Techniques?

Data analysis is often used to get hard answers when there are more than two factors. This is a good job for multivariate statistics because they let you test theory models that look at how factors are linked using real events. Most of the time, these models mix what we already know with new ideas about how things might be connected.

- When experts use more than one way, they get a fuller picture of how things are linked.
- Find out how strong these links are and which way they go.
- To look at how other things might have an effect, you can use tools like cross-tabulations, partial correlations, and multiple regression.
- Learn more about the place where some links take place.

Multivariate techniques are better than basic techniques for getting deeper ideas and running better statistical tests. This makes it possible to read difficult information in a way that is more true to life.

Challenges in Using Multivariate Techniques

Multivariate methods are useful, but they are hard to understand in a school setting, and you usually need expensive statistics software to use them. It also takes a lot of technical knowledge to figure out what the results mean because the methods are based on assumptions that aren't always simple to check.

It's also very bad that they can't get enough info. If there isn't enough data, the study might have big mistakes. This would mean that the results are not as reliable. Things are more likely to be true when the groups are bigger.

You could learn how to use statistical tools, but you need to know a lot about statistics to fully grasp the results and come to the right conclusions.

How to Select the Right Method for Practical Problems?

Before you can choose the best multivariate method, you need to know all the facts. More than one statistics tool can be used to solve many problems in the real world.

- For instance, when you're building a model to help you guess what will happen,

- Begin with an idea from nature or biology.
- Plan how you will study, and then follow through with your goals for the tests.
- Look over the facts and pick out the most important ones.
- Use the right tools, like principal component analysis, stepwise regression, or correlation analysis, to find ties and pick the most important factors.

Set up a way to make guesses and make it better.

Look over the model, make it better, and then use it in the real world. Multivariate analysis problems have a lot of different ways to solve them. Because it gives people a full picture that helps them pick the right path and get things done faster.

Very important parts of multivariate analysis

Model-based analytical methods, principal component analysis (PCA), and multivariate analysis of variance (MANOVA) are some of the best tools for multivariate analysis. And you can only use these methods if you meet certain statistical assumptions, like those about normality, regression, homogeneity of variance, and more. Each of these methods has its own pros and cons. This means that during the study design step, there needs to be careful planning for the app to work. To do this, they need to come up with a theory framework, decide what data to collect, how to collect it, and what research tools to use.

The Linear Model Approach in Multivariate Analysis

Multivariate techniques can be broadly categorized into three main groups:

1. Linear Model Methods:
   - This group includes the tests for multivariate analysis of variance (MANOVA), analysis of covariance, and multiple regression analysis.
   - These ways use linear models to figure out how one or more factors that are not dependent on each other are connected.
2. Classification Techniques:
   - Polytomous analysis and discriminant function analysis are two of them. Polytomous analysis looks at more than one type.
   - Their main goal is to use prediction factors to put the results into groups that have already been decided.
3. Data Reduction Techniques:
   - It has canonical correlation analysis, factor analysis, and principal component analysis in it.
   - For these methods to work, they need to find a smaller group of hidden variables that account for most of the changes in the first dataset. These variables are also called factors or components.

Multivariate Analysis of Variance (MANOVA)

MANOVA decomposes the total variance in a dataset based on its sources—typically defined by the experimental design. It assesses both the main effects of independent variables and the interaction effects among them on one or more dependent variables.

For example, in a 2 × 2 factorial design, total variance is divided into:

- The variance due to the two main factors (each with two levels),
- The interaction variance between the two factors,
- The error variance (within-group variation).

The significance of each component is then evaluated using F-tests.

Advantages of MANOVA include:

- The ability to test multiple factors and levels simultaneously.
- Evaluation of interactions between variables and their combined effects on dependent variables.

Limitations include:

- The requirement that each group consists of independent, randomly selected samples.
- The assumption that data are normally distributed and homoscedastic (equal variance across groups).

Key Multivariate Technique: Multiple Regression Analysis

Multiple regression analysis is a statistical method used to explore the linear relationship between a single dependent variable and multiple independent variables.

The general form of the regression equation is:

$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m + \varepsilon$

Where:

- y is the dependent variable,
- x_1, x_2, ..., x_m are the independent variables,
- α is the intercept,
- β_1 to β_m are the regression coefficients,
- ε is the random error term.

By inputting the values of the independent variables into the regression model, one can predict the corresponding value of the dependent variable. This makes multiple regression a powerful tool for:

- Quantitative description of relationships,
- Prediction and forecasting,
- Application to both continuous and binary (dichotomous) outcome variables.

Linear Regression Techniques

This technique explores the linear relationship between one or more dependent variables (y's) and one or more independent variables (x's). The analysis is based on fitting a linear model that relates the dependent variables to the predictors, with a focus on estimating and testing the parameters of the model. A key consideration is determining which independent variables should be included in the model—especially when the relevant predictors are not known in advance.

Three main cases can be distinguished based on the number of variables involved:

1. Simple Linear Regression:

   Involves one dependent variable (y) and one independent variable (x).

   Example: Predicting a student's college GPA using only their high school GPA.

2. Multiple Linear Regression:

   Involves one dependent variable and multiple independent variables.

   Example: Predicting college GPA based on a combination of factors such as high school GPA, standardized test scores (ACT/SAT), and high school quality ratings.

3. Multivariate Multiple Linear Regression:

   Involves multiple dependent variables and multiple predictors.

   Example: Predicting multiple academic outcomes—such as GPA in science, arts, and humanities, or years of college completed—using a range of predictors.

Logistic regression, often referred to as a "choice model," is a variation of multiple regression that is used when the dependent variable is binary or categorical (e.g., yes/no, success/failure). Unlike linear regression, the goal here is to predict the probability of an event occurring.

- The separate factors can be a single thing or a series of things.
- A contingency table is often used to see how the planned and real groups stack up against each other.
- Model accuracy is measured by how well predicted outcomes match actual results—both correctly predicted occurrences and non-occurrences, divided by the total sample size.

This method is particularly valuable in areas such as consumer behavior research, where it helps predict choices made by individuals when presented with various options.

Multivariate Analysis of Variance (MANOVA)

MANOVA is an extension of ANOVA that examines the effect of categorical independent variables on multiple metric (continuous) dependent variables simultaneously.

- While ANOVA evaluates group differences using T-tests (for two means) or F-tests (for more than two means), MANOVA assesses how a set of dependent variables is influenced by one or more categorical factors.
- Commonly used in experimental designs, MANOVA tests whether the mean vectors of dependent variables differ significantly across groups defined by the categorical variables.

Important considerations include:

- Sample size: Ideally, 15–20 observations per group (cell) are required. Having too many observations per cell (e.g., over 30) may reduce the effectiveness of the analysis.
- Balanced design: Group sizes (cells) should be roughly equal. The largest group should not exceed 1.5 times the size of the smallest group.
- Assumption of normality: The dependent variables should follow a normal distribution.

The effectiveness of the MANOVA model is assessed by comparing group mean vectors. If statistically significant differences are found, the null hypothesis is rejected, indicating that the treatments or groupings have a measurable effect.

Factor Analysis

It's best to keep a study plan as easy as possible by concentrating on just a few main points. When you use factor analysis, you don't need a dependent variable to look at data. It instead looks in the data grid for the pattern that is hiding there. It is thought that the factors in this method will stay the same and stay spread out. There should be three to five variables that put a lot of weight on each factor. There should be at least five counts for each variable, which means there should be fifty records.

This kind of research can be done in two main ways:

- Common Factor Analysis: Focuses on the shared variance among variables and is typically used to uncover unobservable (latent) factors that underlie the observed measurements.
- Principal Component Analysis (PCA): Concentrates on explaining the total variance in the dataset by identifying the fewest possible number of components that account for the maximum variance.

Factors are usually extracted based on specific criteria, such as eigenvalues greater than 1.0, or using a Scree Plot to visually determine the optimal number of factors. In factor analysis, the original observed variables ($y_1, y_2, ..., y_p$) are represented as linear combinations of a smaller set of unobservable variables ($f_1, f_2, ..., f_m$), where $m < p$. These unobserved variables (factors) represent the hidden dimensions that "generate" the observed data. While the factors themselves cannot be directly measured, they vary across individuals in a manner similar to the original variables.

The main goal of factor analysis is to cut down on duplicate data by combining many variables into a smaller number of factors that are easier to understand. That way, the important info stays the same, but the collection is eased out.

Discriminant Function Analysis

(Group Separation Analysis)

Based on a set of predictors, discriminant function analysis (DFA) sorts cases into groups that have already been set. The key idea is to construct discriminant functions using known group membership data, and then apply these functions to assign new cases to the appropriate group.

There are two primary purposes of this analysis:

1. The discriminant function is a list of factors that show how two or more groups are different from each other. This is what "group separation" means. This kind of research helps figure out which variables have the most impact on how groups are different and which projection space best displays how groups are split up.
2. Guessing or Sorting Observations: To guess which group new observations will be in, you can use linear or quadratic classification functions. We figure out which group a case is most likely to belong to by looking at the numbers it gets on the indicator factors.

Although DFA is generally used with continuous variables, it can also be applied to qualitative data using appropriate numerical encoding techniques. This allows researchers to develop objective classification rules based on empirical data. However, discriminant analysis is only applicable when the categories or groups are already known. If group membership is not predetermined, cluster analysis should first be used to identify natural groupings in the data, after which discriminant analysis can be applied to validate or interpret the resulting clusters.

Cluster Analysis

Clue analysis is the process of putting records with more than one variable into groups. This is how you look through a set of data for trends. The primary objective is to form groups (clusters) such that observations within the same cluster are similar to each other, while observations in different clusters are dissimilar. Ideally, the resulting clusters represent natural groupings that are meaningful and interpretable within the context of the research.

Cluster analysis addresses the problem of unsupervised statistical classification. Given a dataset of n objects, each described by p observed variables, the goal is to determine how these objects can be grouped into a number of unknown classes. If the classification is based on similarities between objects, the analysis is referred to as Q-type clustering. If it involves grouping variables, it is known as R-type clustering. The fundamental principle of clustering is to minimize intra-cluster variation and maximize inter-cluster variation.

One common approach is hierarchical clustering, in which the process starts by treating each of the n objects as its own separate cluster. The algorithm then calculates pairwise distances (or dissimilarities) between all clusters, merges the two closest clusters,

and repeats the process iteratively until *k* clusters are formed, as specified by the analyst or derived through statistical criteria.

Unlike traditional classification methods, which require predefined group labels, cluster analysis is exploratory: neither the number of clusters nor their composition is known in advance. Many clustering techniques begin by measuring similarities or distances between all pairs of observations—commonly using metrics such as Euclidean distance. Other methods, such as k-means clustering, begin with initial guesses for cluster centroids and iteratively reassign observations based on proximity to these centers. Some clustering techniques may also rely on minimizing within-cluster variance while maximizing between-cluster variance.

Things and forces can be grouped together. A lot of the time, correlation numbers are used to find links between two things. This is very useful when you want to find trends but don't have any data to use.

Making things easy to see can help with cluster analysis. Maybe a simple scatterplot is all you need if you only have two factors (p = 2). You can use biplots or Principal Component Analysis (PCA) to show data in two ways, even if it has more than two. This shows how the groups are set up.

Another name for cluster analysis is number taxonomy. It is also known as pattern recognition and independent learning. A lot of different areas use it, like engineering, history, marketing, sociology, crime, archaeology, geology, medicine, and marketing.

When working with large datasets, the first thing that is often done is to separate the n events into k groups. First, grouping optimization rules are used to keep making changes until a stable classification that makes sense is found.

Despite its strong statistical foundation, cluster analysis often yields results that depend heavily on the choice of algorithm and distance metrics. Hence, expert judgment is required to validate and interpret the final cluster solution. It is often necessary to compare results from multiple clustering techniques to identify the solution that best aligns with the goals and domain-specific requirements of the study.

The structure of the data can be expressed in matrix form as:

$$Y = \begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{pmatrix} = \begin{pmatrix} y^{(1)} & y^{(2)} & \cdots & y^{(p)} \end{pmatrix}$$

Where y_i' represents an observation vector (a row), and $y^{(j)}$ corresponds to the j-th variable (a column).

Multidimensional Scaling (MDS)

Multidimensional Scaling (MDS) is a statistical technique used to convert subjective judgments of similarity or dissimilarity among items into a spatial representation. The goal is to position items in a multidimensional space such that the perceived similarities

correspond to the distances between them. As a decompositional and exploratory method, MDS is particularly valuable in uncovering latent dimensions of perception and in revealing how individuals compare products when the specific basis for comparison is not explicitly defined.

MDS is classified as a **dimensionality reduction** technique. It begins with a set of dissimilarities or distances δ_{ij} between pairs of *n* items. The objective is to construct a spatial configuration of the items in a lower-dimensional space such that the Euclidean distances d_{ij} between items closely approximate the original dissimilarities δ_{ij}, i.e., $d_{ij} \approx \delta_{ij}$ for all item pairs.

These dissimilarities may originate from two sources:

- **Objective measures**: When actual values of observations are available in a p-dimensional space (e.g., y_i and y_j), dissimilarity is often calculated using Euclidean distance:

$$\delta_{ij} = \frac{1}{2}(y_i - y_j)'(y_i - y_j)$$

- **Subjective judgments**: When the input is based on perceived similarity (e.g., consumer ratings of how similar different brands are), δ_{ij} reflects psychological or perceptual distances rather than physical ones.

The primary output of MDS is a perceptual map, a graphical display that illustrates the relationships among the items. This visualization can aid in interpretation by highlighting clusters or relative positions. In some cases, the configuration may lie close to a curve in a two-dimensional space, allowing for the ranking (seriation) of items along that curve.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique designed to reduce the dimensionality of a dataset by transforming the original variables into a new set of uncorrelated variables—called principal components—that sequentially maximize variance. The first principal component captures the largest possible variance in the data, and each subsequent component captures the maximum remaining variance under the constraint of being orthogonal to the preceding components.

T he method is useful when there is a need to rank or compare observations using a single scale derived from multiple variables. For example, if students are assessed across multiple subjects—such as English, mathematics, and reading—PCA can generate a composite score that more effectively differentiates among students than a simple average, by assigning optimal (and potentially unequal) weights to each subject.

Unlike techniques such as regression, canonical correlation, or discriminant analysis—where variables are divided into dependent and independent sets or observations are grouped—PCA treats all variables equally and assumes no predefined groupings. It is fundamentally a one-sample technique applied to datasets where all p variables are considered simultaneously to reveal their internal structure.

PCA also plays a foundational role in other multivariate methods:

- It can serve as a preliminary step in factor analysis, identifying major components that explain the variance before interpreting latent constructs.
- It can assist in identifying influential variables in predictive modeling by highlighting those contributing most to variance in the dataset.

However, PCA focuses exclusively on the internal relationships among variables and does not explore relationships between two distinct sets of variables. For such purposes—like studying the interaction between independent and dependent variables—canonical correlation analysis is more appropriate.

Correspondence Analysis

Correspondence Analysis (CA) is a dimensionality reduction technique used to create a perceptual map based on object ratings across various attributes. Similar to Multidimensional Scaling (MDS), it analyzes both independent and dependent variables simultaneously. However, CA is more closely related to factor analysis in terms of methodology. It is a compositional method, particularly useful when dealing with a large number of attributes and entities (such as brands or companies), especially when those attributes are too similar for factor analysis to yield meaningful results.

The core of CA lies in the construction and analysis of a contingency table, which summarizes the frequency counts of two categorical variables. The goodness of fit for the model is assessed using the Chi-square statistic, making it possible to evaluate the significance of relationships between rows and columns.

Interpretation of the resulting dimensions can be complex, as the axes represent combinations of both independent and dependent variables. The output is a graphical representation where both row and column profiles are displayed in a joint low-dimensional space, helping to visualize the associations and interactions between categories.

Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) is an extension of multiple correlation analysis and serves as a bridge between multivariate regression and correlation methods. It examines the relationship between two sets of continuous variables, identifying linear combinations—called canonical variates—that are maximally correlated.

The procedure involves deriving pairs of canonical variables from each variable set. The first pair exhibits the highest possible correlation between the two sets, followed by successive pairs with decreasing correlations, each uncorrelated with previous pairs. The canonical correlation coefficients quantify the strength of association between the canonical variates.

CCA is most appropriate when both variable sets follow a multivariate normal distribution and when the goal is to understand the shared variance or underlying relationship between the two sets, rather than predicting one from the other.

Conjoint Analysis

Conjoint Analysis, often referred to as trade-off analysis, is a technique used to evaluate preferences for products or services based on their attributes and corresponding levels. It is both a compositional and dependence technique, as it determines how individual components contribute to overall preference.

Each level of each attribute is assigned a part-worth utility (or value), and the total preference for a product configuration is calculated by summing the part-worths of its components. This way can help people find the best blend of traits and values for their needs. A lot of people use it to make things, set prices, and divide markets into different groups.

Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM) is a group of methods that let you find many links between unknown and known factors at the same time. There are measurement models and structure models in SEM. Measurement models show how clear and hidden variables are connected, while structure models show how hidden variables are connected to each other. Plenty of other words only talk about straight lines, but this one does more.

Though we can't see hidden traits like intelligence, happiness, or drive, we can get a good idea of them from things we can measure, like polls or test scores. SEM is a strong tool that is often used to test theory models, especially ones with factors that can change the results. People in psychology, schooling, and the social studies can use it.

Log-Linear Models

Log-linear models are use statistics to figure out how the groups in a contingency table are linked to each other. These models don't really tell the difference between factors that depend on variables and variables that depend on factors. Instead, they look at how all the parts are connected at the same time.

You can use a chosen log-linear model to figure out how often each number in the table of possibilities is likely to occur. After that, statistics like are used to see how the numbers that were expected matched up with the ones that actually happened.

- Pearson's Chi-square (χ^2)
- Likelihood Ratio Statistic (G^2)

If the calculated values of χ^2 and G^2 are less than the corresponding critical values from the Chi-square distribution table, the model is considered to fit the data well. Otherwise, alternative models must be tested to better represent the observed data structure.

We collected our data from Al-Rafidain General Hospital, one of the largest hospitals in Iraq specializing in cardiovascular diseases. The sample size included N = 980 patients. In this study, we applied the **Log-linear model** technique for data analysis.

The patients were categorized based on the following variables:

1. Type of Cardiovascular Disease (considered here as the dependent variable) with two categories:

A – Coronary artery disease

B – Hypertension

2. Age groups (in years): divided into three categories:
3. A – 30 to 49

B – 50 to 69

C – 70 and above

Patients younger than 30 years old were excluded because, according to hospital records, cardiovascular incidents in that age group were extremely rare.

The data were arranged in a two-way contingency table as follows:

Disease	**30-49**	**50-69**	**70+**
Coronary artery disease	120	310	110
Hypertension	140	280	120

Next, we calculated the expected frequencies under the assumption of independence between the two variables using the formula:

$$\frac{(_{j.}x)(_{.i}x)}{N} = {}_{ij}m$$

where $_{.i}x$ and $_{j.}x$ represent the marginal totals for the i-th row and j-th column respectively, and N is the total sample size.

The expected values matrix was computed as follows:

+70	**50-69**	**30-49**	**Disease**
110.0	305.7	124.3	Coronary artery disease
120.0	284.3	135.7	Hypertension

Following that, we performed goodness-of-fit tests using the **Pearson Chi-square** ($^2\chi$) and the **Likelihood Ratio test** (G^2), computed by:

$$^2\chi = \sum \frac{^2(_{ij}x_{ij} - m)}{_{ij}m}$$

$$^2G = 2\sum x_{ij} \ln\left(\frac{_{ij}x}{_{ij}m}\right)$$

:The calculated statistics were

$$2.345 = {}^2\chi$$

$$2.112 = {}^2G$$

:The degrees of freedom for this two-way table under the independence model are

$$2 = (1-3)(1-2) = (1-c)(1-r) = df$$

At a significance level $= \alpha 0.05$, the critical ${}^2\chi$ value is 5.99. Since both ${}^2\chi$ and ${}^2G$ statistics are less than the critical value, we fail to reject the null hypothesis, indicating no significant association between age .groups and type of cardiovascular disease among the sample studied

Conclusion

We need to use multivariate data analysis right now because it lets us see many things that are linked at the same time. Studies that only look at one part of the data are less useful than studies that use this method on all kinds of data. Methods like factor analysis, cluster analysis, discriminant analysis, and principal component analysis are great for reducing the amount of data you have, putting it into groups, and showing how things are connected. When you use multivariate methods, you should make sure to follow some rules and choose the right ones for your data and study goals. Now that computers and statistical tools are better, it's easier to use these hard studies. They are used more often in engineering, medicine, the social sciences, and medicine because of this. At the end of the day, multivariate data analysis helps you make better predictions, find deeper trends, and use that data to make decisions. It does make a big difference in how science goes forward and how problems are fixed in the real world.

References

Bryant, F. B., & Yarnold, P. R. (1994). Principal components analysis and exploratory and confirmatory factor analysis. *American Psychological Association Books*. ISBN 978-1-55798-273-5.

Byrne, B.M. (2013). Structural equation modeling with AMOS: Basic concepts, applications, and programming, second edition. *Structural Equation Modeling with Amos Basic Concepts Applications and Programming Second Edition*, 1-396, https://doi.org/10.4324/9780203805534

Cai, T. Tony (2013). Sparse PCA: Optimal rates and adaptive estimation. *Annals of Statistics*, *41*(6), 3074-3110, ISSN 0090-5364, https://doi.org/10.1214/13-AOS1178

Chandra, S., & Menezes, D. (2001). Applications of Multivariate Analysis in International Tourism Research: The Marketing Strategy Perspective of NTOs. *Journal of Economic and Social Research*, 3(1), 77–98.

Dong, C. (2014). Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. *Accident Analysis and Prevention*, *70*, 320-329, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2014.04.018

Feinstein, A. R. (1996). *Multivariable Analysis*. Yale University Press.

Garson, D. (2009). Factor Analysis. *Statnotes: Topics in Multivariate Analysis*. Retrieved April 13, from http://www2.chass.ncsu.edu/garson/pa765/statnote.htm

Ge, T. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1), ISSN 2041-1723, https://doi.org/10.1038/s41467-019-09718-5

Hahs-Vaughn, D.L. (2016). Applied multivariate statistical concepts. *Applied Multivariate Statistical Concepts*, 1-647, https://doi.org/10.4324/9781315816685

Huque, M.H. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies 01 Mathematical Sciences. *BMC Medical Research Methodology*, *18*(1), ISSN 1471-2288, https://doi.org/10.1186/s12874-018-0615-6

Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal*, *39*(8), 2362-2385, ISSN 0143-2095, https://doi.org/10.1002/smj.2783

Kononen, D.W. (2011). Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. *Accident Analysis and Prevention*, *43*(1), 112-122, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2010.07.018

Mix, K.S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology General*, *145*(9), 1206-1227, ISSN 0096-3445, https://doi.org/10.1037/xge0000182

Morais, C.L.M. (2020). Tutorial: multivariate classification for vibrational spectroscopy in biological samples. *Nature Protocols*, *15*(7), 2143-2162, ISSN 1754-2189, https://doi.org/10.1038/s41596-020-0322-8

Obozinski, G. (2011). Support union recovery in high-dimensional multivariate regression. *Annals of Statistics*, *39*(1), 1-47, ISSN 0090-5364, https://doi.org/10.1214/09-AOS776

Olive, D.J. (2017). Linear regression. *Linear Regression*, 1-494, https://doi.org/10.1007/978-3-319-55252-1

Olivieri, A.C. (2015). Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. *Analytica Chimica Acta*, *868*, 10-22, ISSN 0003-2670, https://doi.org/10.1016/j.aca.2015.01.017

Rahmati, O. (2020). Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Science of the Total Environment*, *699*, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2019.134230

Raubenheimer, J. E. (2004). An item selection procedure to maximize scale reliability and validity. *South African Journal of Industrial Psychology*, 30(4), 59–64.

Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd ed.). John Wiley & Sons, Inc.

Riaz, A. (2011). Role of the EASL, RECIST, and WHO response guidelines alone or in combination for hepatocellular carcinoma: Radiologic-pathologic correlation. *Journal of Hepatology*, *54*(4), 695-704, ISSN 0168-8278, https://doi.org/10.1016/j.jhep.2010.10.004

Shi-Sheng, D. (1981). *Multiple Analysis Method and Its Applications*. Jilin People's Publishing House.

Tsuruoka, Y., Tsujii, J., & Ananiadou, S. (2009). Gradient Descent Training for L1-Regularized Log-linear Models with Cumulative Penalty. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP* (pp. 477–485). Suntec, Singapore.

Yao Ting, Z., & Kaitai, F. (1982). *Multivariate Statistical Analysis Introduction*. Science Press.

Yu, P. (2018). Design of experiments and regression modelling in food flavour and sensory analysis: A review. *Trends in Food Science and Technology*, *71*, 202-215, ISSN 0924-2244, https://doi.org/10.1016/j.tifs.2017.11.013