Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1611775

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 8 - Issue 2: 428-434 / March 2025

APPLYING PERMANOVA FOR MULTIVARIATE ANALYSIS OF VARIANCE IN HEALTH STUDIES

Fahrettin KAYA[1]*

[1]Kahramanmaraş Sütçü İmam University, Andırın Vocational School, Department of Computer Technologies, 46400, Kahramanmaraş, Türkiye

Abstract: Health data often do not meet the normality assumption, which limits the applicability of traditional analysis of variance methods. The aim of this study is to propose a methodological framework for analyzing such data by examining PERMANOVA (Permutational Multivariate Analysis of Variance), a method that does not require the normality assumption and is particularly suitable for complex datasets, within the context of maternal health data. In the context of maternal healthcare in Bangladesh, the effects of two independent variables—risk and age factors—on multivariate response variables such as Systolic Blood Pressure (mmHg), Diastolic Blood Pressure (mmHg), Blood Sugar (mmol/L), Body Temperature (Fahrenheit), and Heart Rate (beats per minute) were examined using the PERMANOVA method. The first independent variable represents the risk factor, comprising three different risk levels (low, mid, high), while the second independent variable represents the age factor, divided into four age groups (young, adolescent, middle-aged, menopausal). The dependent variables did not follow a normal distribution, as confirmed by the Anderson-Darling test and Mardia's multivariate normality test. As a result of the PERMANOVA analysis, it was determined that at least two mean differences between the groups of the risk factor and the age factor were statistically significant in terms of the response variables ($P<0.01$). Furthermore, pairwise comparisons of the factor groups revealed that the mean differences between the low, mid, and high levels of the risk factor, as well as the mean differences among the young, adolescent, and middle-aged groups of the age factor, were statistically significant ($P<0.01$). However, the mean difference between the middle-aged and menopausal groups for the age factor was found to be statistically insignificant ($P>0.01$). The PERMANOVA method is recommended for researchers to accurately determine whether the mean differences in factor levels are statistically significant or to identify threshold values of the groups by using multiple response variables simultaneously and performing pairwise comparisons of factor groups.

Keywords: PERMANOVA, Pairwise comparison, ANOVA, Non-parametric analysis, Maternal health

***Corresponding author:** Kahramanmaraş Sütçü İmam University, Andırın Vocational School, Department of Computer Technologies, 46400, Kahramanmaraş, Türkiye
E mail: fkaya@ksu.edu.tr (F. KAYA)
Fahrettin KAYA https://orcid.org/0000-0003-1666-4859

Received: January 02, 2025
Accepted: February 02, 2025
Published: March 15, 2025

Cite as: Kaya F. 2025. Using permanova for multivariate analysis of variance in health research: An application with R. BSJ Eng Sci, 8(2): 428-434.

1. Introduction

Health is one of the most fundamental aspects of human life, and research in this field generally focuses on the causes, effects, and treatments of diseases. For instance, diabetes can lead to complications such as diabetic neuropathy and nephropathy, and it may also contribute to the development of cardiovascular diseases and other vascular disorders (Sowers et al., 2001). The development of treatment methods to address such complex health problems represents a significant area of research. The effectiveness of these treatment methods across patient groups should also be investigated.

In such studies, statistical methods provide powerful tools for understanding complex relationships. For example, Analysis of Variance (ANOVA) is a frequently used method in health research. ANOVA is a fundamental statistical tool used to examine variance differences in the means of a dependent variable across different groups of an independent variable (Underwood, 1981; Şahin and Koç, 2019). However, ANOVA is limited to analyzing single dependent variables. In cases in which multiple dependent variables are involved, more comprehensive methods, such as multivariate analysis of variance (MANOVA) are often preferred (Tinsley and Brown, 2000; Tabachnick et al., 2013; Koç et al., 2019). Nevertheless, MANOVA methods may encounter difficulties in meeting key assumptions such as multivariate normality and homogeneity.

In this context, Permutational Multivariate Analysis of Variance (PERMANOVA) has emerged as a nonparametric statistical test method for analyzing complex multivariate datasets (Anderson, 2001). Due to its flexibility and versatility, PERMANOVA is widely used in various fields, including biology, ecology, and environmental sciences. The proposed method allows multivariate analysis independent of data distribution. For instance, using this method, Pasin et al. (2016) demonstrated that individuals with Hashimoto's disease exhibit significantly different lipid levels compared with healthy individuals, regardless of sex.

Considering these challenges, this study makes a methodological contribution by applying the nonparametric multivariate analysis technique PERMANOVA to maternal health data. The study examined the association between risk factors and age groups (independent variables) and key health indicators such as systolic and diastolic blood pressure, blood glucose levels, body temperature and heart rate (dependent variables) and analyzed the mean differences between the factor groups in terms of dependent variables. The findings are expected to provide valuable information for maternal health risk assessments and contribute to the widespread use of nonparametric methods in medical research.

2. Materials and Methods

The dataset used in this study was created by collecting data from healthcare centers, transferring it via an IoT smart health device to a web portal, and storing it on a server. The dataset was obtained from a machine learning repository published on the Kaggle platform and includes 1,014 patient records (Ahmed and Kashem, 2020; Togunwa et al., 2023).

The first independent variable in the dataset is the patient risk level. This variable was categorized into three groups: high risk (272 records), medium risk (336 records), or low risk (406 records). The second independent variable is age group. This variable was restructured according to women's fertility status and categorized into four categories. The young age group (under 20 years) had low fertility levels, with 279 records. The adolescent age group (20-34 years) had the highest fertility levels, with 417 records. The middle age group (35-44 years) exhibits declining fertility levels and includes 135 records. The menopausal age group (above 44 years) had low fertility levels and comprised 183 records. The dependent variables examined in this study comprised five key health-related parameters: diastolic blood pressure (DBP, mmHg), body temperature (BT, °F), systolic blood pressure (SBP, mmHg), blood glucose (BG, mmol/L), and heart rate (HR, bpm). The dataset was analyzed to explore the interactions between these factors and independent variables to evaluate health outcomes. The dataset was analyzed using PERMANOVA to explore the interactions between factors and dependent variables.

PERMANOVA analysis is frequently employed in data structures that exhibit MANOVA patterns. Evaluations of differences between groups were conducted via a transformation predicated on ANOVA. A crucial facet of this analysis is transformation. The F value in the ANOVA table was not derived directly from the raw data; instead, it was calculated on the basis of a distance matrix constructed from the distances between pairs of observations. In ANOVA or MANOVA, the original layout of the raw data was grouped by factors, as depicted in Figure 1a. The grouping order is randomly shuffled hundreds of times, rearrangements are executed (permutations are performed), and pseudo-F values are calculated each time over the distance matrices created by these permutations. Concurrently, the F value of the original layout wass recorded in the initial calculation. A permutation test statistic is obtained by comparing the F-value in the original order with the pseudo F-values, and the difference between the groups is scrutinized with this test statistic (Anderson, 2001; 2014). The distances referenced in the study typically were included distance measurements such as Euclidean, City Block, Chi-square, Bray Curtis, Binomial, Jaccard, and others (Oksanen et al., 2022). In this study, the Bray Curtis, Binomial, and Mahalanobis distance measures were favored. The assumption in this analysis does not necessitate uniformity of variance, provided that a balanced experimental design is in place. The observations (rows in the original data matrix) are deemed independent and originate from an identical distribution.

2.1. Data Preparation

The starting point of this analysis was the creation of a distance observation matrix using the raw data matrix. In the raw data matrix, each observation is represented by a row in the original order of the factor groups, and each response variable is represented by a column. This matrix consists of N rows and p columns and is represented as seen in Figure1a (Anderson, 2001).

The matrix of dimensions N x N, comprising distances between row pairs of observations with indices such as i and j, undergoes transformation as depicted in Figure1b. The matrix derived from this transformation is termed the observation distance matrix. Each cell within the matrix cells the distance value between pairs of response variables. This matrix exhibits a symmetrical structure, composed of gray and white regions. The SS_T is computed utilizing the observation distance matrix in Figure1a's grey area.

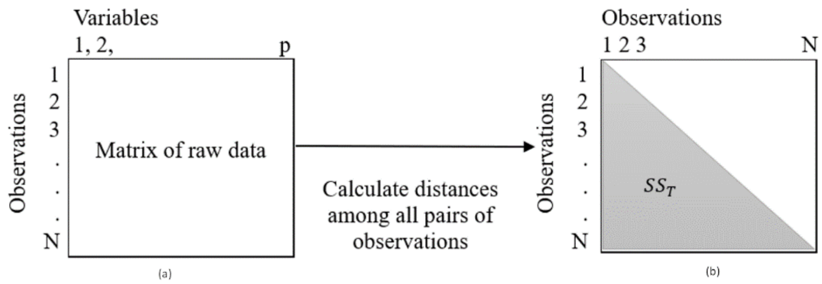

Figure 1. Illustrative representation of the data structure in PERMANOVA analysis.

The distance between the observation rows i and j is denoted by d_{ij}. For instance, the Euclidean distance of three-response variables such as X, Y, Z is calculated by ANOVA by reducing it to a singular distance value as $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$. In this scenario, the squares of all distances in the lower-diagonal half of the distance matrix are aggregated (total variance) and divided by the total number of observations N. The sum of the resulting grand total squares is as follows in equation 1:

$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \quad (1)$$

2.2. One-Way ANOVA Design

In the instance of a one-way design in the PERMANOVA analysis, in the most elementary scenario, consider a groups in which a factor such as A is examined, and n observations in each group. The total number of observation units N equals an. In this scenario, the exemplar design, which comprises two groups such as single-factor A_2 and A_1, can be depicted as shown in Figure 2a. The summation of the within squares of factor A corresponds to the summation of the distance values d_{ij} in the cells constituting the gray area in Figure 2a. That is, $SS_{W(A)}$ symbolizes the summation of the squared distances between repetitions within the same group and is computed by dividing by n, which represents the number of repetitions of each group. The summation of intragroup squares is expressed as in equation 2:

$$SS_{W(A)} = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2\, \delta_{ij} \quad (2)$$

here, the value of δ_{ij} is 1 when observation i and observation j are in the same group; otherwise, its value is zero. In addition, the total squares between the groups of factor A are given by in equation 3.

$$SS_A = SS_T - SS_{W(A)} \quad (3)$$

Using these calculations, a one-way analysis of variance table is arranged as shown in Appendix A.

2.3. Two-way ANOVA Design

Consider a two-factorial design scenario with factors A (e.g., Risk Factor) and B (e.g., Age Factor) in the context of statistical analysis. Factor A, the primary factor, comprises ‘a’ levels or groups, each containing ‘n’ observations. Similarly, Factor B, the secondary factor, also consists of ‘b’ distinct levels or groups, each with ‘n’ observations. The total number of observations, denoted by N, can be mathematically represented as $N = abn$. The distance matrices of observations in a two-factorial design can be depicted as illustrated in Figure 2 (Anderson, 2001).

When the effect of Factor B is disregarded, the within-group sum of squares for Factor A, denoted as $SS_{W(A)}$, is obtained by dividing the sum of squared distances in the cells falling into the grey area in Figure 2a by ‘bn’, and it is calculated using equation 4.

$$SS_{W(A)} = \frac{1}{bn} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2\, \delta_{ij}^A \quad (4)$$

here, δ_{ij}^A takes the value of 1 if observation i and observation j are in the same group of factor A; otherwise, it takes the value of zero. In addition, the sum of squares between the groups of Factor A, denoted as SS_A, is obtained by the equation $SS_A = SS_T - SSW(A)$. Similarly, disregarding the effect of Factor A, the within-group sum of squares for Factor B, denoted as $SS_{W(B)}$, is calculated by dividing the sum of squared distances in the cells falling into the grey area in Figure 2b by the number of observations (‘an’), and it is calculated using equation 5:

$$SS_{W(B)} = \frac{1}{an} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2\, \delta_{ij}^B \quad (5)$$

here, δ_{ij}^B takes the value of 1 if observation i and observation j are in the same group of Factor B; otherwise, it takes the value of zero.

In a two-factorial PERMANOVA analysis, the third sum of squares represents the sum of squares of residuals. This total, denoted as S_R, is obtained by dividing the sum of the squared distances in the cells falling into the gray area in Figure 2c by the number of observations “abn”. That is, the inter-point distances are calculated for each combination of 'ab' for Factors A and B, as shown in equation 6.

$$SS_R = \frac{1}{abn} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2\, \delta_{ij}^{AB} \quad (6)$$

here, δ_{ij}^{AB} takes the value of 1 if observation i and observation j are in the same combination of Factors A and B; otherwise, it takes the value of zero. If there is an interaction between the two factors in the design, the sum of squares related to this interaction, denoted as SS_{AB}, is calculated, as shown in equation 7.

$$SS_{AB} = SS_T - SS_A - SS_B - SS_R \quad (7)$$

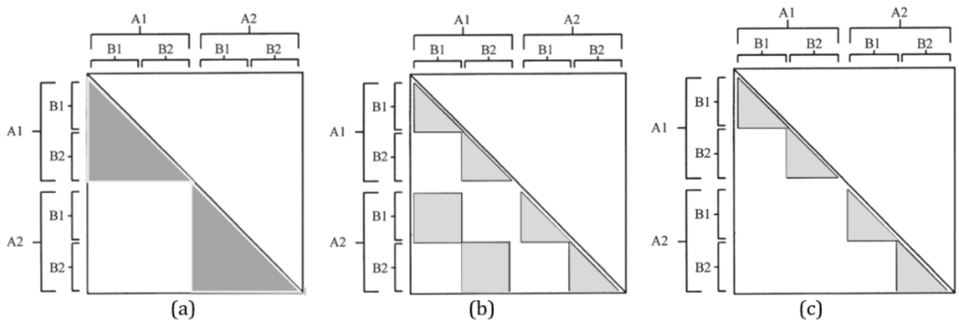

Figure 2. Representation of single or two-factor observation distance matrix shape.

If there is no interaction term in the design, that is, if Factor B is within Factor A, the sum of squares for Factor B within A, denoted as $SS_{B(A)}$, is calculated directly as follows as in equation 8:

$$SS_{B(A)} = SS_T - SS_A - SS_R \quad (8)$$

Following these calculations, the results of the two-way PERMANOVA analysis are organized as shown in Appendix B. In the computation of the table, the 'adonis2' function from the 'vegan' library (Oksanen et al., 2022), which is integrated into the R software, was utilized and coded as follows:

```
library(vegan)
library(readxl)
mydata <- read_excel("C:/Users/…/mydata.xlsx")
mydata<-data.frame(mydata)
mydata$RiskLevel<-as.factor(mydata$RiskLevel)
mydata$AgeGrp<-as.factor(mydata$AgeGrp)
responses = as.matrix(mydata[,3:7])
result <- adonis2(responses ~ RiskLevel*AgeGrp,
            data = mydata, nperm= 999, method="bray")
result
AIC_BIC_value<- calculate_AIC_BIC(result)
AIC_BIC_value
```

In the source code, the dataset in Excel format was assigned to the 'mydata' variable, and the risk and age variables were converted to factors. At the same time, response variables 'SystolicBP', 'DiastolicBP', 'BS', 'BodyTemp', and 'HeartRate' were converted into a matrix and assigned to the 'responses' variable. This transformation was written to obtain the distance observation matrix in Figure1b. In the analysis, a function with arguments 'adonis2(response ~ RiskLevel * AgeGrp, data = data, nperm = 999, method = "bray")' was used. In addition, 999 permutations were performed on the observation distance matrix created using the Bray Curtis, Binomial, and Mahalanobis distance measures.

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values and R-square (explained variance ratio) were used to select the best model among these three models. Lower AIC and BIC values indicate better model performance, while the R-square value measures the fit of the model to the data (Wikipedia, 2023). Since the Adonis2 function cannot calculate the AIC and BIC values directly, a function called calculate_AIC_BIC was developed to calculate these values. This function takes the result variable, which is the output of the Adonis2 function, as the input and calculates the AIC and BIC values. The calculated AIC and BIC values are assigned to the variable AIC_BIC_value and made ready for use. The details of this function are presented in Appendix C.

When a statistically significant p-value is obtained from the ANOVA analysis, multiple comparison tests are performed to determine from which groups the mean difference between factor groups originates. Paired comparisons of factor groups were made in pairs using the PERMANOVA method. The "RVAideMemoire" library integrated into the R software was used to perform this operation (Hervé ,2022). The "pairwise.perm.manova" function of this library has been coded in R programing as follows:

```
library(RVAideMemoire)
responsesDMatrix<-vegdist(responses, method="bray")
pairwiseAge <-pairwise.perm.manova(
            responsesDMatrix, mydata$AgeGrpp,
             p.method="holm", nperm=999,F=TRUE)
pairwiseAge
pairwiseRisk <-pairwise.perm.manova(
            responsesDMatrix, mydata$RiskLevel,
method="holm", nperm=999,F=TRUE)
pairwiseRisk
```

In the source code, the matrix of response variables calculated using Bray Curtis distance was transferred to the variable 'responsesDmatrix'. The results of the pairwise comparison calculations made for the age factor were assigned to the 'pairwiseAge' variable, and the result of the pairwise comparison calculations made for the risk factor was assigned to the 'pairwiseRisk' variable. The 'pairwise.perm.manova' function was used for both factors, and the arguments were identified as 999 for the number of permutations, Bray Curtis, Binomial, and Mahalanobis for the distance measures, and the Holm method to control for type 1 error. The distributions of the response variables were investigated for compliance with the normality assumption. According to the results of Mardia's multivariate normality test (Korkmaz et al., 2014). It was determined that the SystolicBP, DiastolicBP, BS, BodyTemp, and HeartRate variables do not have a multivariate normal distribution, and this was calculated using the R code as follows:

```
library(MVN)
result <- mvn(responses, mvnTest = "mardia")
print(result)
```

The source code generates both multivariate normality test (Mardia's multivariate normality test) and univariate normality test (Anderson-Darling test) results with the parameter mvnTest = "mardia". The obtained results are assigned to the result variable.

3. Results

Within the scope of this research, the effects of risk and age factors on response variables such as, 'SystolicBP', 'DiastolicBP', 'BS', 'BodyTemp', and 'HeartRate' were examined. The Mardia Skewness and Kurtosis statistics for the multivariate normality test of the response variables were presented, and normality tests were performed separately for each response variable using the Anderson-Darling method. The results are presented in Table 1.

Skewness and kurtosis statistics revealed significant deviations from multivariate normality ($P<0.01$) for all response variables ('SystolicBP', 'DiastolicBP', 'BS', 'BodyTemp', and 'HeartRate').

Table 1. Multivariate and univariate normality test table

		Multivariate	
Test	Statistic	P value	Result
Mardia Skewness	1619.85	<0.001	NO
Mardia Kurtosis	18.55	0	NO
		Univariate	
Variable	F Statistic	P value	Normality
SystolicBP	50.76	<0.001	NO
DiastolicBP	18.38	<0.001	NO
BS	139.50	<0.001	NO
BodyTemp	242.30	<0.001	NO
HeartRate	16.88	<0.001	NO

Table 2. Table of PERMANOVA analysis results

Variable	Df	SumOfSqs	R^2	F	Pr(>F)	AIC	BIC
		Mahalanobis distance					
RiskLevel	2	551.4	0.11	69.78	0.001	4279	4338
AgeGrp	3	337.3	0.07	28.45	0.001		
RiskLevel:AgeGrp	6	216.6	0.04	9.14	0.001		
Residual	1002	3959.6	0.78				
Toplam	1013	5065.0	1				
		Binomial distance					
RiskLevel	2	0.75	0.38	411.33	0.001	-4206	-4147
AgeGrp	3	0.26	0.13	97.056	0.001		
RiskLevel:AgeGrp	6	0.05	0.03	9.58	0.001		
Residual	1002	0.91	0.46				
Toplam	1013	1.98	1				
		Bray Curtis distince					
RiskLevel	2	0.44	0.15	107.77	0.001	-3398	-3339
AgeGrp	3	0.33	0.11	54.24	0.001		
RiskLevel: AgeGrp	6	0.13	0.04	10.27	0.001		
Residuals	1002	2.03	0.70				
Totals	1013	2.91	1				

Df= degree of freedom, R^2= coefficient of determination, AIC= Akaike Information Criterion, BIC= Bayesian Information Criterion.

Indicating they did not follow a normal distribution. Given this non-normality, PERMANOVA, a non-parametric methodology, was employed to analyze the impact of risk level (RiskLevel) and age group (AgeGrp) variables on response variables, as presented in Table 2.

The PERMANOVA analysis, using Mahalanobis, Binomial, and Bray-Curtis distances, revealed a significant impact of both risk and age factors on the response variables ($p < 0.01$). This means there's a significant difference between the means of at least two subgroups within each factor. While higher R-squared values are generally good indicators of model fit and reliability, they can also suggest overfitting, especially when paired with low residuals. In this case, the PERMANOVA analysis with the binomial distance measure showed a residual variance of 46%, meaning that risk and age factors explain 54% of the total variance. Other factors not included in the model likely contribute to the remaining unexplained variance. Additionally, this model has low AIC (-4206) and BIC (-4147) values. Therefore, the binomial distance is preferred for further analysis. The next step involved using PERMANOVA to compare the means of the factor subgroups/levels, as shown in Table 3.

Table 3 reveals statistically significant mean differences between the sublevels of both the risk and age factors, except for the comparison between the middle-aged (35-44) and menopausal (44+) groups (P>0.05). This implies that an expectant mother's risk level and age group, except for being middle-aged or menopausal, significantly influence the response variables. Furthermore, while the p-values were similar across all three distance measures, the PERMANOVA analysis with binomial distance exhibited the highest R-squared value, indicative of a stronger relationship between the model's dependent and independent variables.

R version 4.3.1 was used in this study (R Core Team, 2023).

4. Discussion and Conclusion

The health data of pregnant women in Bangladesh were analyzed using the PERMANOVA method to evaluate the effects of maternal risk status and age on response variables. Table 1 reveals that the response variables did not satisfy the assumption of normality.

Table 3. Factor groups' paired mean comparison table

Risk and Age Groups	Bray Curtis		Binomial		Mahalanobis	
	R^2	P	R^2	P	R^2	P
Low risk - Mid risk	0.19	0.003	0.41	0.003	0.13	0.003
Low risk-High risk	0.12	0.003	0.36	0.003	0.09	0.003
Mid risk-High risk	0.03	0.003	0.03	0.003	0.03	0.003
Young age (<20)- Adolescent age (20-34)	0.10	0.006	0.09	0.006	0.05	0.006
Young age (<20)- Middle age (35-44)	0.21	0.006	0.30	0.006	0.09	0.006
Young age (<20)- Menopause (44>)	0.27	0.006	0.34	0.006	0.11	0.006
Adolescent age (20-34)- Middle age (35-44)	0.04	0.006	0.21	0.006	0.04	0.006
Adolescent age (20-34)- Menopause (44>)	0.07	0.006	0.26	0.006	0.06	0.006
Middle age (35-44)- Menopause (44>)	<0.0001	0.698	<0.0001	0.989	<0.0001	0.935

R^2= coefficient of determination, Bray Curtis= Bray Curtis distance, Binomial= Binomial distance, Mahalanobis= Mahalanobis distance.

PERMANOVA utilized distances such as Mahalanobis, Binomial, and Bray-Curtis, enabling the examination of relationships between factor groups from different perspectives. According to the analysis results, the PERMANOVA model based on the binomial distance demonstrated good performance in the multivariate analysis, with a high R-squared value and low AIC and BIC values. Furthermore, the results of this model indicate that age and risk factors have a statistically significant effect on the response variables ($P<0.01$). Ahmed and Kashem (2020) classified risk groups with 97% accuracy using machine learning methods on the same dataset. The accuracy rate provides further evidence of significant differences in response variables across risk groups.

The literature supports the applicability of the PERMANOVA method to various fields. For instance, Nascimento et al. (2019) used this method to identify ecological differences between groups and found significant differences in pairwise comparisons of factor groups.

In this study, age was also converted into a categorical format and categorized into four groups. According to the pairwise comparison results in Table 3, no significant difference was found between the "Middle Age (35-44)" and "Menopause (44+)" groups ($p>0.01$). This finding provides valuable insights for researchers when determining threshold values for age groups.

In conclusion, the PERMANOVA method is a robust and flexible tool for evaluating the effects of factors on response variables in health research, especially when the normality assumption is not met. In addition, this method provides a framework for researchers to determine threshold values for factor groups.

Author Contributions

The authors' contributions are detailed below. The authors have reviewed and approved the final version of the manuscript.

	F.K.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Approval/Informed Consent

The datasets used in this study are publicly accessible and can be found at the following link: https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data

Acknowledgments

We sincerely thank Ahmed and Kashem for making the open-access dataset used in this study available.

References

Ahmed M, Kashem MA. 2020. IoT-based risk level prediction model for maternal health care in the context of Bangladesh. 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), December 19-20, Dhaka, Bangladesh, pp: 1-6. https://doi.org/10.1109/STI50764.2020.9350320

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol, 26(1): 32-46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Anderson MJ. 2014. Permutational multivariate analysis of variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat07841

Hervé M. 2022. RVAideMemoire: Testing and plotting procedures for biostatistics [Computer software]. CRAN. https://CRAN.R-project.org/package=RVAideMemoire

Koç Ş, Çanga Boğa D, Önem AB, Yavuz E, Şahin M. 2019. Monte Carlo simulation study robustness of MANOVA test statistics in Bernoulli and uniform distribution. BSJ Eng Sci, 2(2): 42-51. https://doi.org/10.34248/bsengineering.563330

Korkmaz S, Göksülük D, Zararsiz G. 2014. MVN: An R package for assessing multivariate normality. The R J, 6(2): 151-162. https://doi.org/10.32614/RJ-2014-031

Nascimento FJA, Dahl M, Deyanova D, Lyimo LD, Bik HM, Schuelke T, Bonaglia S. 2019. Above-below surface interactions mediate effects of seagrass disturbance on meiobenthic diversity, nematode and polychaete trophic structure. Commun Biol, 2: 362. https://doi.org/10.1038/s42003-019-0607-z

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2022. Vegan: Community ecology package [Computer software]. CRAN. https://CRAN.R-project.org/package=vegan

Pasin O, Ankarali H, Cangür S, Sungur MA. 2016. Nonparametric multivariate analysis of variance and its application in the field of health. J Info Technol, 9(1): 13-20.

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Sowers JR, Epstein M, Frohlich ED. 2001. Diabetes, hypertension, and cardiovascular disease: An update. Hypertension, 37(4): 1053-1059. https://doi.org/10.1161/01.HYP.37.4.1053

Şahin M, Koç Ş. 2018. A Monte Carlo simulation study robustness of MANOVA test statistics in Bernoulli distribution. Süleyman Demirel Univ J Inst Nat App Sci, 22(3): 1125-1131.

Tabachnick BG, Fidell LS, Ullman JB. 2013. Using multivariate statistics. Pearson, London, UK, pp: 983 (6th ed.).

Tinsley HEA, Brown SD. 2000. Multivariate statistics and mathematical modeling. In Handbook of applied multivariate statistics and mathematical modeling. Academic Press, Cambridge, US, pp: 3–36. https://doi.org/10.1016/B978-012691360-6/50002-8

Togunwa TO, Babatunde AO, Abdullah KU. 2023. Deep hybrid model for maternal health risk classification in pregnancy: Synergy of ANN and random forest. Front AI, 6: 1213436. https://doi.org/10.3389/frai.2023.1213436

Underwood AJ. 1981. Techniques of analysis of variance in experimental marine biology and ecology. Oceanog Marine Biol: An Annual Rev, 19(Suppl 1): 513-605.

Wikipedia. 2023. Akaike information criterion. URL: https://en.wikipedia.org/wiki/Akaike_information_criterion https://en.wikipedia.org/wiki/Bayesian_information_criterion (accessed date: September 30, 2023).