Atom Indonesia Vol. 49 No. 3 (2023) 215 - 231

Atom Indonesia

Journal homepage: http://aij.batan.go.id

Temporal Trends and Spatial Relationships of Radioactive Isotopes (I-131, Cs-134, and Cs-137) in Response to Nuclear Events: *A Comprehensive Analysis Using Time Series Graphs, Regression, and Multivariate Techniques*

B. Nasution[1*], W. Ritonga[2], A. Doyan[3], P. D. Pandara[4], L. Alfaris[5], R. C. Siagian[6]

[1,2,6]*Physics Departemen, Universitas Negeri Medan, Medan 20221, Indonesia*
[3]*Master of Science Education, Postgraduate Program, University of Mataram, Mataram 83115, Indonesia*
[4]*Department of Physics, Sam Ratulangi University, Manado 95115, Indonesia*
[5]*Department of Marine Technology, Politeknik Kelautan dan Perikanan Pangandaran, Pangandaran 395008, Indonesia*

A R T I C L E I N F O

Article history:
Received 17 February 2023
Received in revised form 3 November 2023
Accepted 3 November 2023

Keywords:
Radioactive isotopes
Evolution
Longitude and latitude
Linear regression
Principal component analysis (PCA)

A B S T R A C T

This research aims to comprehend the evolution of radioactive isotopes Iodine-131 (I-131), Cesium-134 (Cs-134), and Cesium-137 (Cs-137) over time in diverse locations and analyze their relationships with the independent variables Longitude and Latitude using Linear Regression, Principal Component Analysis (PCA), and Canonical Correlation Analysis (CCA). The data used in this study were processed from the "DE.xlsx" file, including the imputation of missing values with 0 and column transformation into factors. The results of the Linear Regression analysis indicate a significant association between these isotopes and Longitude and Latitude. Additionally, PCA and CCA analyses reveal complex relationships between the isotopes and independent variables. This research provides valuable insights into the historical trends of radioactive isotopes Iodine-131 (I-131), Cesium-134 (Cs-134), and Cesium-137 (Cs-137) in various locations. The novel aspect and uniqueness of this study lie in the utilization of a comprehensive analytical approach, combining Linear Regression, PCA, and CCA to comprehend the relationships between isotopes and specific environmental factors. Moreover, this study significantly contributes to understanding the phenomena of radioactive isotopes and can serve as a foundation for further research in this field. The findings of this research are expected to support efforts in preventing and managing potential environmental and human health impacts of radioactive isotopes in the future.

INTRODUCTION

Background

The background of this research is related to the analysis of radioactive isotopes Iodine-131 (I-131), Cesium-134 (Cs-134), and Cesium-137 (Cs-137) data at various locations. These isotopes are generally formed as a result of nuclear reactions involving nuclear reactors, nuclear explosions, or nuclear accidents [1]. The presence of these isotopes in the environment can have impacts on both the environment and human health [2]. This study relies on processed data from the "DE.xlsx" file, involving critical steps like filling NAs with 0, converting columns to numeric data, and transforming others into factors. This rigorous data preparation ensures the data's integrity. Subsequently, the processed data is employed to generate time series plots for various radioactive isotopes. These plots visually depict the evolving radioactivity trends, enabling the identification of patterns and changes across

*Corresponding author.
E-mail address: budimannasution@unimed.ac.id
DOI: https://doi.org/10.55981/aij.2023.1306

different locations [3]. The research results offer valuable insights into the historical trends of radioactive isotopes I-131, Cs-134, and Cs-137, aiding our understanding of radioactivity changes, their environmental and health impacts, and enhancing nuclear activity management through precise data processing techniques for reliable findings.

This research aims to analyze the radioactivity levels of I-131, Cs-134, and Cs-137 at different locations, exploring their connection with Longitude and Latitude. It offers insights into isotope behavior over time, identifies trends, and assesses the impact of nuclear events on environmental radioactivity [4]. Enhancing awareness of statistical analysis methods like linear regression and hypothesis testing to empower researchers and stakeholders in addressing radioactivity risks effectively.

This research exclusively examines three radioactive isotopes (I-131, Cs-134, and Cs-137) utilizing data from the "DE.xlsx" file with some processing. Notably, the data's origin and collection location are unspecified. Regression analysis was exclusively performed on I-131 and Cs-137 using Longitude and Latitude as independent variables. Additional statistical analyses, including the Chi-Square test and Factor Analysis, were exclusively conducted on I-131, Cs-134, and Cs-137 isotopes.

Research findings on radioactive isotope behavior offer crucial insights for nuclear security, environmental health, and policy-making, enabling effective preventive and mitigation measures.

This research innovatively utilizes radioactive isotopes I-131, Cs-134, and Cs-137 from diverse locations, employing various statistical analyses to reveal relationships with Longitude and Latitude. It also involves regression analysis and hypothesis testing. While yielding valuable results, future studies can improve by explaining data sources and location origins, enhancing statistical analysis, exploring other isotopes, and examining additional influencing factors. This novel approach provides insights into radioactive isotope behavior and management, fostering future research opportunities.

In this research, Fig. 1 illustrates the geographical coordinates used to measure radiation exposure from the Chernobyl nuclear accident. Figure 2 focuses on Location Group 1 in Germany, the starting point for analyzing radiation pollution. Figure 3 visually represents time series analysis of Iodine-131, Cesium-134, and Cesium-137 using R-Studio. Figure 4 uses R-Studio for regression analysis of these isotopes in relation to Longitude and Latitude. Figures 5 to 22 display various analyses conducted with R-Studio, offering deep insights into radiation data and its influencing factors in specific regions, contributing to our scientific understanding of isotopic radiation impact.

This article comprehensively analyzes the behavior of isotopes I-131, Cs-134, and Cs-137 in connection to nuclear incidents, citing the work of Assimakopoulos, Ioannides, and Paradopoulou [5] on their transport into cheese products post-Chernobyl. Hashimoto, S., Komatsu, M., and Miura's [6] article, 'Radioactive Materials Released by the Fukushima Nuclear Accident,' delves into the release of radioactive materials during the Fukushima nuclear disaster. Article by Masson et al. [7] discusses radioiodine releases in nuclear emergencies.

METHODOLOGY

Coordinate and latitude processing

In the context of the Chernobyl accident, isotope pollution refers to the release of radioactive materials into the environment due to a leak at the Chernobyl nuclear reactor in Ukraine in 1986 [8]. This pollution resulted in the spread of radioactive substances, including certain isotopes, such as I-131, Cs-134, and Cs-137. It is possible to analyze the Mean Longitude data of these countries to understand the extent of the impact of Chernobyl isotope pollution on geographic data variation across countries, with a focus on the distribution of Mean Longitude.

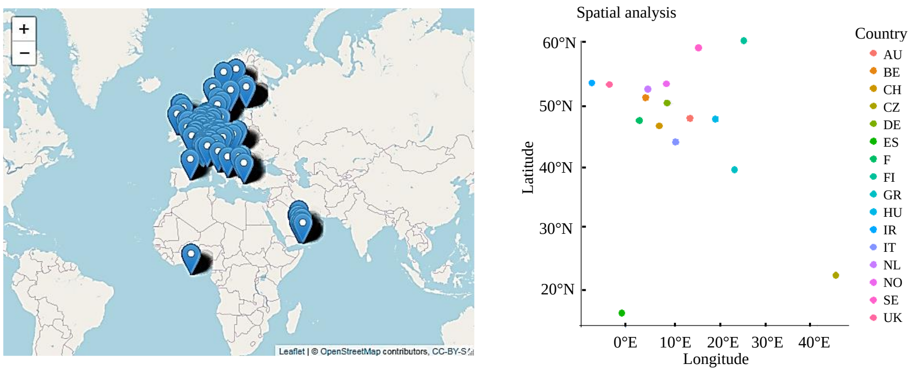


Fig. 1. Longitude and latitude coordinates of isotope radiation exposure measured due to the Chernobyl nuclear accident.

Figure 1 illustrates that some countries, like Ireland (IR) and the United Kingdom (UK), exhibit a symmetrical distribution of Mean Longitude with the median slightly below zero, indicating a specific trend. In contrast, countries such as Spain (ES) and Norway (NO) display a more variable Mean Longitude distribution, with a wider range and outlying values, suggesting a higher degree of variation. These variations may reflect the impact of isotope pollution from the Chernobyl accident on environmental and ecological conditions in different regions [9].

Table 1 provides key statistics on Latitude and Longitude coordinates for different countries, highlighting significant variability in their geographic positioning. For example, the Czech Republic (CZ) exhibits notably diverse spatial data points with high standard deviations, while countries like Hungary (HU) and Iran (IR) have data concentrated at a single location, indicated by zero standard deviations. This table aids in evaluating the dispersion and central tendencies of geographic data in these countries.

Table 1. Statistics on the mean and standard deviation of Latitude and Longitude coordinates of each country.

PAYS	Mean Longitude	SD Longitude	Mean Latitude	SD Latitude
AU	13.9	1.99	47.7	0.638
BE	4.71	0.398	51	0.201
CH	7.52	0.489	46.4	0.123
CZ	43.8	11.4	22.3	13.4
DE	9.22	2.13	50.2	1.79
ES	-0.164	0.199	16.2	19.6
F	3.51	2.17	47.3	2.19
FI	24.9	0.132	60.2	0.141
GR	23	0.806	39.4	1.24
HU	19.1	0	47.5	0
IR	-6.28	0	53.4	0
IT	10.9	2.59	43.8	2
NL	5.22	1.03	52.4	0.622
NO	8.96	3.19	53.3	18.8
SE	15.6	3.05	59.1	2.83
UK	-2.63	1.4	53.1	1.94

Fig. 2. Selection of group 1 location will concentrate attention on the analysis and exploration of pollution in Germany.

Figure 2 highlights the significance of selecting Germany as Group 1's focal point, allowing an in-depth analysis of pollution using provided data. The research aims to uncover isotope value variations in I-131, Cs-134, and Cs-137 across different German locations, establishing connections between geographic coordinates and pollution levels. This involves applying diverse statistical analysis methods for valuable insights.

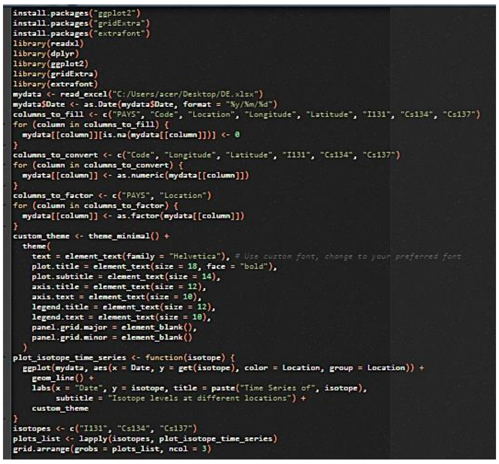
Time series graphs

This research employed time series analysis to track the trends of radioactive isotopes, including I-131, Cs-134, and Cs-137, using data from the 'DE.xlsx' file. Preprocessing steps addressed missing values, data type conversion, and column transformation to enhance subsequent analysis.

After processing the data, time series plots were created for radioactive isotopes I-131, Cs-134, and Cs-137, revealing historical trends and behavior. These insights are vital for understanding environmental and human health implications, enhancing the study's accuracy and effectiveness.

Such research is vital for tracking and analyzing radiation levels at various sites to evaluate the consequences of nuclear incidents, tests, or other factors that may raise radioactive isotope levels in the environment. This knowledge empowers researchers and policymakers to make informed choices about radioactive material safety, ecosystem impact, and human population well-being.

In Fig. 3, the study employed R-Studio programming language for time series analysis, offering a comprehensive exploration of Iodine-131 (I131), Cesium-134 (Cs134), and Cesium-137 (Cs137) isotopes' temporal evolution, revealing valuable insights into their behavior and trends.



```r
install.packages("ggplot2")
install.packages("gridExtra")
install.packages("extrafont")
library(readxl)
library(dplyr)
library(ggplot2)
library(gridExtra)
library(extrafont)
mydata <- read_excel("C:/Users/acer/Desktop/DE.xlsx")
mydata$Date <- as.Date(mydata$Date, format = "%y/%m/%d")
columns_to_fill <- c("PAYS", "Code", "Location", "Longitude", "Latitude", "I131", "Cs134", "Cs137")
for (column in columns_to_fill) {
  mydata[[column]][is.na(mydata[[column]])] <- 0
}
columns_to_convert <- c("Code", "Longitude", "Latitude", "I131", "Cs134", "Cs137")
for (column in columns_to_convert) {
  mydata[[column]] <- as.numeric(mydata[[column]])
}
columns_to_factor <- c("PAYS", "Location")
for (column in columns_to_factor) {
  mydata[[column]] <- as.factor(mydata[[column]])
}
custom_theme <- theme_minimal() +
  theme(
    text = element_text(family = "Helvetica"), # Use custom font, change to your preferred font
    plot.title = element_text(size = 18, face = "bold"),
    plot.subtitle = element_text(size = 14),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  )
plot_isotope_time_series <- function(isotope) {
  ggplot(mydata, aes(x = Date, y = get(isotope), color = Location, group = Location)) +
    geom_line() +
    labs(x = "Date", y = isotope, title = paste("Time Series of", isotope),
         subtitle = "Isotope levels at different locations") +
    custom_theme
}
isotopes <- c("I131", "Cs134", "Cs137")
plots_list <- lapply(isotopes, plot_isotope_time_series)
grid.arrange(grobs = plots_list, ncol = 3)
```

Fig. 3. R-Studio programming language for time series analysis methods to investigate and describe the evolution of certain radioactive isotopes, namely Iodine-131 (I131), Cesium-134 (Cs134), and Cesium-137 (Cs137) over time.

Regression analysis

Regression analysis is a statistical technique used to investigate the relationship between a dependent variable (also known as the response variable) and one or more independent variables (also known as predictor variables) [10]. It helps to understand how changes in the independent variables are associated with changes in the dependent variable. The analysis uses simple linear regression. Simple linear regression is a type of regression analysis that examines the linear relationship between one independent variable and one dependent variable [11]. It fits a straight line to the data points to represent the relationship. Multiple linear regression is an extension of simple linear regression, and it allows us to investigate the relationship between one dependent variable and multiple independent variables [12]. It fits a linear equation to the data points, considering the combined effects of all the independent variables on the dependent variable.

```
library(stats)
perform_regression <- function(isotope) {
  formula <- as.formula(paste0(isotope, " ~ Longitude + Latitude"))
  model <- lm(formula, data = mydata)
  return(summary(model))
}
isotopes <- c("I131", "Cs134", "Cs137")
for (iso in isotopes) {
  print(paste("Regression Analysis for", iso))
  print(perform_regression(iso))
}
```

Fig. 4. R-Studio programming language for regression analysis performed on different isotopes (I-131, Cs134, and Cs137) with respect to their relationship with two independent variables, longitude and latitude.

In Fig. 4, the research employed the R-Studio programming language to conduct a regression analysis on various isotopes, namely I131, Cs134, and Cs137. The analysis aimed to explore the relationship between these isotopes and two independent variables, Longitude and Latitude coordinates. For the simple linear regression analysis, the mathematical model for each isotope can be represented as Eqs. (1-3).

For I131:

$$Y_{\text{I-131}} = \beta_0^{\text{I-131}} + \beta_1^{\text{I-131}} \cdot X_1 \tag{1}$$

For Cs134:

$$Y_{\text{Cs134}} = \beta_0^{\text{Cs134}} + \beta_1^{\text{Cs134}} \cdot X_1 \tag{2}$$

For Cs137:

$$Y_{\text{Cs137}} = \beta_0^{\text{Cs137}} + \beta_1^{\text{Cs137}} \cdot X_1 \tag{3}$$

For the multiple linear regression analysis, the mathematical model for each isotope can be represented as Eqs. (4-6).

For I131:

$$Y_{\text{I-131}} = \beta_0^{\text{I-131}} + \beta_1^{\text{I-131}} \cdot X_1 + \beta_2^{\text{I-131}} \cdot X_2 \tag{4}$$

For Cs134:

$$Y_{\text{Cs134}} = \beta_0^{\text{Cs134}} + \beta_1^{\text{Cs134}} \cdot X_1 + \beta_2^{\text{Cs134}} \cdot X_2 \tag{5}$$

For Cs137:

$$Y_{\text{Cs137}} = \beta_0^{\text{Cs137}} + \beta_1^{\text{Cs137}} \cdot X_1 + \beta_2^{\text{Cs137}} \cdot X_2 \tag{6}$$

In these equations, β_0 represents the intercept, and β_1 and β_2 represent the coefficients of the independent variables (Longitude and Latitude) for each isotope.

Descriptive analysis

This research method is descriptive analysis. Descriptive analysis is a research method aimed to systematically and concisely describe and present data [13]. This study utilizes descriptive analysis to describe the characteristics of three radioactive isotopes, namely I-131, Cs-134, and Cs-137, along with several related independent variables. In descriptive analysis Fig. 5, the researcher presents relevant descriptive statistics for each isotope, including the mean, median, mode, standard deviation, variance, minimum value, and maximum value [14]. Additionally, the researcher also calculates the first quartile (Q1) and third quartile (Q3) to depict the data distribution.

```
isotope_columns <- c("I131", "Cs134", "Cs137")
calculate_descriptive_statistics <- function(data) {
  desc_stats <- data.frame(
    Mean = mean(data, na.rm = TRUE),
    Median = median(data, na.rm = TRUE),
    Mode = as.numeric(names(which.max(table(data)))),
    Standard_Deviation = sd(data, na.rm = TRUE),
    Variance = var(data, na.rm = TRUE),
    Minimum = min(data, na.rm = TRUE),
    Maximum = max(data, na.rm = TRUE),
    Q1 = quantile(data, probs = 0.25, na.rm = TRUE),
    Q3 = quantile(data, probs = 0.75, na.rm = TRUE)
  )
  return(desc_stats)
}
for (isotope in isotope_columns) {
  cat("Descriptive Statistics for", isotope, ":\n")
  stats_result <- calculate_descriptive_statistics(mydata[[isotope]])
  print(stats_result)
  cat("\n")
}
```

Fig. 5. R-Studio programming language for descriptive analysis.

One sample t-test

The research used the One Sample t-test method to test hypotheses about the means of three isotopes: I-131, Cs-134, and Cs-137. The t-test is a statistical method used to determine if there is a significant difference between the mean of a sample and a hypothesized value [15]. The following is a step-by-step explanation of the One Sample t-test method used in the study.

For null Hypothesis (H_0), the true mean of each isotope (I-131, Cs-134, Cs-137) is equal to zero (μ = 0). For alternative Hypothesis (Ha), the true mean of each isotope (I-131, Cs-134, Cs-137) is not equal to zero ($\mu \neq 0$). The study collected samples of each isotope (I-131, Cs-134, Cs-137) and calculated their sample means (mean of x). The t-statistic is a measure of how many standard errors the sample mean is away from the hypothesized mean (μ = 0) and is calculated using the formula, see Eq. (7) [16].

$$t = \frac{\text{mean of } x - \text{hypothesized mean}}{\text{standard error of the mean}} \quad (7)$$

The degrees of freedom represent the number of independent pieces of information in the sample data used to calculate the t-statistic [17]. It is calculated as the sample size minus 1 (df = sample size - 1) [17]. The p-value is the probability of obtaining results as extreme or more extreme than the observed results, assuming that the null hypothesis is true [18]. A smaller p-value suggests stronger evidence against the null hypothesis [19]. It is compared to a significance level (usually 0.05) to determine statistical significance. The 95 % confidence interval provides a range of values within which the true mean of the population (μ) is likely to lie with 95 % confidence [20].

Let $\bar{x}_{I131}$, $\bar{x}_{Cs134}$, and $\bar{x}_{Cs137}$ represent the sample means for isotopes I-131, Cs-134, and Cs-137, respectively. Let s_{I131}, s_{Cs134}, and s_{Cs137} represent the standard deviations of the samples for isotopes I-131, Cs-134, and Cs-137, respectively. The standard error (SE) is calculated as Eqs. (8-10).

$$SE_{I131} = \frac{s_{I131}}{\sqrt{n}}, \quad (8)$$

where n is the sample size for isotope I-131:

$$SE_{Cs134} = \frac{s_{Cs134}}{\sqrt{n}}, \quad (9)$$

where n is the sample size for isotope Cs-134:

$$SE_{Cs137} = \frac{s_{Cs137}}{\sqrt{n}}, \quad (10)$$

where n is the sample size for isotope Cs-137.

The t-statistic measures how many standard errors the sample mean is away from the hypothesized mean $(\mu = 0)$ and is calculated as Eqs. (11-13).

$$t_{I131} = \frac{\bar{x}_{I131} - 0}{SE_{I131}}, \quad (11)$$

$$t_{Cs134} = \frac{\bar{x}_{Cs134} - 0}{SE_{Cs134}}, \quad (12)$$

$$t_{Cs137} = \frac{\bar{x}_{Cs137} - 0}{SE_{Cs137}}. \quad (13)$$

The degrees of freedom represent the number of independent pieces of information in the sample data used to calculate the t-statistic and is calculated as the sample size minus 1, see Eq. (14).

$$df = n - 1, \quad (14)$$

where n is the sample size for each isotope. The p-value is the probability of obtaining results as extreme or more extreme than the observed results, assuming that the null hypothesis is true [18]. It can be calculated using a t-distribution table or statistical software based on the t-statistic and degrees of freedom [21]. If the p-value is smaller than the chosen significance level (usually 0.05), we reject the null hypothesis in favor of the alternative hypothesis, indicating a statistically significant difference in the mean of the isotope [22]. The 95 % confidence interval provides a range of values within which the true mean of the population (μ) is likely to lie with 95 % confidence [20]. The confidence interval can be calculated using the t-distribution and is given Eqs. (15,16).

$$CI_{lower} = \bar{x} - (t_{critical} \times SE), \quad (15)$$

$$CI_{upper} = \bar{x} + (t_{critical} \times SE), \quad (16)$$

where $\bar{x}$ is the sample mean, SE is the standard error of the mean, and $t_{critical}$ is the critical value of the t-distribution corresponding to the chosen confidence level and degrees of freedom [23].

Correlation coefficients

The Pearson correlation coefficient measures the level of linear correlation between two continuous variables [24]. It indicates how closely the data points align to a straight line, representing the degree of linear association [25]. The formula to calculate the Pearson correlation coefficient between two variables X and Y is as Eq. (17).

$$r_{xy} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \quad (17)$$

where r_{xy} is the Pearson correlation coefficient between X and Y. X_i and Y_i are individual data points of X and Y, respectively. $\overline{X}$ and $\overline{Y}$ are the means of X and Y, respectively. The Pearson correlation coefficient ranges from -1 to 1, in which [26]: A value of 1 indicates a perfect positive linear relationship; A value of -1 indicates a perfect negative linear relationship; A value of 0 indicates no linear relationship between the variables.

Meanwhile the Spearman correlation coefficient, also known as rank correlation, assesses the strength and direction of a monotonic relationship between two variables [27]. Monotonic relationships imply that as one variable increases, the other variable either increases or decreases, but not necessarily at a constant rate [28]. To compute the Spearman correlation coefficient between variables X and Y, the data is first converted into ranks, and then the Pearson correlation formula is applied to the ranked data [29]. The formula for the Spearman correlation coefficient between X and Y is similar to the Pearson correlation coefficient formula, but it uses the ranks of the data, see Eq. (18).

$$\rho_{xy} = \frac{\sum (R_{X_i} - \overline{R_X})(R_{Y_i} - \overline{R_Y})}{\sqrt{\sum (R_{X_i} - \overline{R_X})^2 \sum (R_{Y_i} - \overline{R_Y})^2}} \quad (18)$$

where ρ_{xy} is the Spearman correlation coefficient between X and Y. R_{X_i} and R_{Y_i} are the ranks of the individual data points of X and Y, respectively. $\overline{R_X}$ and $\overline{R_Y}$ are the means of the ranks of X and Y, respectively. The Spearman correlation coefficient also ranges from -1 to 1, with similar interpretations as the Pearson coefficient.

As shown in Fig. 6, this research employed both Pearson and Spearman correlation coefficients to assess the relationships between the concentrations of I-131, Cs-134, and Cs-137. The Pearson coefficient measured linear relationships, while the Spearman coefficient measured monotonic relationships based on ranked data [30]. The correlation coefficients were interpreted to determine the strength and direction of the relationships between the variables [11].

```r
library(Hmisc)
cor_columns <- c("I131", "Cs134", "Cs137")
cor_pearson <- rcorr(as.matrix(mydata[cor_columns]), type = "pearson")
cor_spearman <- rcorr(as.matrix(mydata[cor_columns]), type = "spearman")
print("Pearson Correlation Coefficients:")
print(cor_pearson$r)
print("\nSpearman Correlation Coefficients:")
print(cor_spearman$r)
```

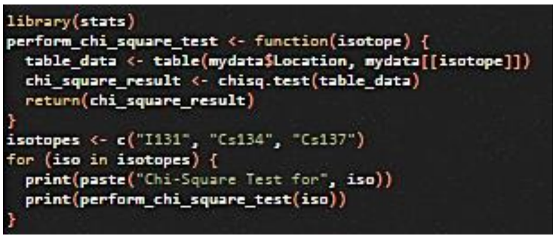
Fig. 6. R-Studio programming language for pearson and spearman coefficient correlation analysis.

Chi-Square methods

The Chi-Square test is a statistical method used to evaluate whether there is a significant association or relationship between two categorical variables in a given dataset [31]. It is widely used in research to analyze the independence or dependence of variables and can be applied to various fields, such as biology, social sciences, and more [32].

The objective of the Chi-Square test is to determine if there is a significant association between two categorical variables in the dataset [33]. Set up null and alternative hypotheses to be tested. The null hypothesis (H_0) assumes that there is no association between the two variables, while the alternative hypothesis (H_1) assumes that there is a significant association between them [34]. Create a contingency table to display categorical variable combinations' frequencies. Calculate expected frequencies assuming no association (null hypothesis). These represent expected frequencies with no variable relationship. The Chi-Square statistic (χ^2) is calculated by comparing the observed frequencies in the contingency table to the corresponding expected frequencies [35]. It measures the discrepancy between the observed and expected frequencies. The degrees of freedom in the Chi-Square test depend on the dimensions of the contingency table. For a 2x2 table, df = 1; but for larger tables, it is calculated as $(\text{rows} - 1) \times (\text{columns} - 1)$. With the Chi-Square statistic and degrees of freedom, we can determine the critical value from the Chi-Square distribution table or use it to calculate the p-value [36]. The p-value represents the probability of obtaining results as extreme or more extreme than what was observed, assuming that the null hypothesis is true [18]. By comparing the p-value to a chosen significance level (commonly 0.05), we can decide whether to reject the null hypothesis or not [18]. If the p-value is smaller than the significance level, we reject the null hypothesis and conclude that there is a significant relationship between the variables [34]. If the p-value is larger than the significance level, we fail to reject the null hypothesis, indicating that there is no significant association [37].

In Fig. 7, the Chi-Square test is a valuable tool for analyzing the relationship between categorical variables in research [35]. It helps researchers determine whether there is a significant association between the variables under investigation. Proper interpretation of the test results and consideration of data limitations and assumptions are crucial for drawing meaningful conclusions from the Chi-Square test [38].



```
library(stats)
perform_chi_square_test <- function(isotope) {
  table_data <- table(mydata$Location, mydata[[isotope]])
  chi_square_result <- chisq.test(table_data)
  return(chi_square_result)
}
isotopes <- c("I131", "Cs134", "Cs137")
for (iso in isotopes) {
  print(paste("Chi-Square Test for", iso))
  print(perform_chi_square_test(iso))
}
```

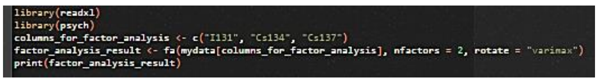
Fig. 7. R-Studio programming language for Chi-Square statistical analysis.

Factor analysis method

Factor Analysis is a statistical method used to identify underlying patterns or factors that explain the associations between variables in a dataset [39]. This technique helps reduce the dimensionality of data and identify the main factors that can account for the variability in observed variables [40]. The methodology used in this study can be summarized as follows.

The study involved three variables, namely I-131, Cs-134, and Cs-137. Measurements were taken to determine the associations between these variables. The data used in Factor Analysis is often standardized to ensure that variables have the same scale [41]. In this case, the standardized loadings, h_2, and u_2 values were used to analyze the relationships between factors and variables [42]. Standardized loadings are standardized coefficients that indicate the strength of the relationship between variables and underlying factors [43]. For each variable, standardized loadings were obtained with respect to each factor. h_2 represents the common variance explained by variables with respect to factors, while u_2 represents the unique variance not explained by variables [44]. These values provide insights into the extent to which variables explain the variability in factors.

Common variance (com) is the total shared variance explained by factors [39]. For each factor, the common variance was calculated [39]. The "minres" method was employed to identify two main factors (MR_1 and MR_2) underlying the associations between the observed variables [45]. MR_1 and MR_2 have different factor loadings, h_2, and u_2 values. The model was evaluated using various goodness-of-fit measures, including objective function, chi-square value, root mean square of residuals (RMSR), and Tucker Lewis Index of factoring reliability [46]. This condition is shown in Fig. 8.



```
library(readxl)
library(psych)
columns_for_factor_analysis <- c("I131", "Cs134", "Cs137")
factor_analysis_result <- fa(mydata[columns_for_factor_analysis], nfactors = 2, rotate = "varimax")
print(factor_analysis_result)
```

Fig. 8. R-Studio programming language for factor analysis.

MR_1 and MR_2 is the two main factors underlying the associations between the observed variables. $h2_{i1}, h2_{i2}, h2_{i3}, h2_{i4}, h2_{i5}, h2_{i6}$ is standardized loadings for variables I-131, Cs-134, and Cs-137 with respect to MR_1 and MR_2. $u2_{i1}, u2_{i2}, u2_{i3}, u2_{i4}, u2_{i5}, u2_{i6}$ is unique variances for variables I-131, Cs-134, and Cs-137 not explained by MR_1 and MR_2. com_{MR_1} and com_{MR_2} is common variance explained by MR_1 and MR_2, respectively. For each variable i (in this case, I-131, Cs-134, and Cs-137) with respect to each factor (MR_1 and MR_2), we have standardized loadings: $h2_{i1}$ is standardized loading for variable i with respect to MR_1; $h2_{i2}$ is standardized loading for variable i with respect to MR_2. For each factor (MR_1 and MR_2), we calculate the common variance as the sum of squared standardized loadings for the variables associated with that factor, see Eqs. (19,20) [45].

$$com_{MR1} = h2^2_{I131} + h2^2_{Cs134} + h2^2_{Cs137} \quad (19)$$

$$com_{MR2} = h2^2_{I131} + h2^2_{Cs134} + h2^2_{Cs137} \quad (20)$$

For each variable i (in this case, I-131, Cs-134, and Cs-137), we calculate the unique variance as the difference between 1 and the sum of the squared standardized loadings for that variable with respect to both factors, See Eqs. (21-23).

$$u2_{I131} = 1 - (h2^2_{I131} + h2^2_{I131}) \quad (21)$$

$$u2_{Cs134} = 1 - (h2^2_{Cs134} + h2^2_{Cs134}) \quad (22)$$

$$u2_{Cs137} = 1 - (h2^2_{Cs137} + h2^2_{Cs137}) \quad (23)$$

Objective Function is the function used to optimize the factor analysis process [47]. Chi-square value is a measure of the discrepancy between the observed and expected covariance matrices [48]. Root Mean Square of Residuals (RMSR) is a measure of the differences between the observed and reproduced correlation matrix [49]. Tucker Lewis Index of Factoring Reliability is a measure of how well the factor analysis model reproduces the input data's covariance matrix [50].

Principal component analysis (PCA) and canonical correlation analysis (CCA) method

In this research, two mathematical models, namely Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA), were utilized to analyze data related to three dimensions (Dim.1, Dim.2, and Dim.3) in a specific context [51]. Eigenvalues were a crucial measure used in these analyses.

PCA is a technique used to reduce the dimensionality of a dataset while retaining as much relevant information as possible [52]. It generates new orthogonal components called principal components that capture the maximum variance in the data.

PCA is a technique used to reduce the dimensionality of data while preserving the most important information [53]. The study's main objective is to discover principal components, such as Dim.1, Dim.2, and Dim.3, which are linear combinations of the original variables, effectively explaining the most variance in the data, as determined by the Eigenvalues analysis [54].

CCA is a statistical method used to explore the relationships between two sets of variables [55]. The goal is to identify optimal linear combinations of two variable sets (isotopes and independent variables) with the highest correlation, particularly focusing on dimensions Dim.1, Dim.2, and Dim.3 in CCA analysis. As shown in Fig. 9, this study evaluates their roles in elucidating the relationship between isotope I-131 and the independent variables, Longitude and Latitude.

```
library(FactoMineR)
library(factoextra)
library(CCA)
library(car)
isotope_data <- mydata[, c("I131", "Cs134", "Cs137")]
pca_result <- PCA(isotope_data, graph = FALSE)
print(pca_result)
dep_vars <- as.matrix(isotope_data)
indep_vars <- as.matrix(mydata[, c("Longitude", "Latitude")])
cca_result <- cc(dep_vars, indep_vars)
print(cca_result)
```

Fig. 9. R-Studio programming language for principal component analysis (PCA) and canonical correlation analysis (CCA).

In this research, PCA was utilized to analyze data related to three dimensions (Dim.1, Dim.2, and Dim.3) in a specific context. It is used to reduce the dimensionality of a dataset while retaining as much relevant information as possible [53].

PCA starts with data preprocessing to standardize the data [56]. Each variable in the dataset is transformed to have zero mean and unit variance. The standardization formula for a variable x_{ij} is as Eq. (24).

$$x_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{24}$$

where x_{ij} is the value of the j-th variable in the i-th data point, μ_j is the mean of the j-th variable, and σ_j is the standard deviation of the j-th variable.

After standardization, the covariance matrix C is calculated [57]. The covariance matrix represents the relationships between variables and is defined as Eq. (25).

$$C = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot x_i^T \tag{25}$$

where n is the number of data points, and x_i is the standardized data vector for the i-th data point.

Next, PCA performs eigenvalue decomposition on the covariance matrix C to find its eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ and corresponding eigenvectors $(v_1, v_2, \ldots, v_m)$. PCA then sorts the eigenvalues in descending order and selects the top k eigenvectors associated with the k largest eigenvalues [58]. These eigenvectors are the principal components that capture the maximum variance in the data. The original data is projected onto the selected k principal components to obtain the reduced-dimensional representation [59]. For the i-th data point, the reduced representation z_i is calculated as Eq. (26).

$$z_i = \begin{bmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{ik} \end{bmatrix} = x_i \cdot V_k \tag{26}$$

where z_i is the k-dimensional vector representing the i-th data point in the reduced space, x_i is the original data point, and V_k is the matrix containing the selected k eigenvectors.

In this research, Canonical Correlation Analysis (CCA) was utilized to explore the relationships between two sets of variables: isotopes (Dim.1, Dim.2, Dim.3) and independent variables (Longitude and Latitude). Similar to PCA, CCA starts with data preprocessing to standardize both sets of variables (X and Y) to have zero mean and unit variance [60]. CCA then calculates the cross-covariance matrix C_{xy} between the two standardized datasets X and Y. The cross-covariance matrix is defined as Eq. (27).

$$C_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i^T \tag{27}$$

where n is the number of data points, x_i is the standardized isotopes vector for the i-th data point, and y_i is the standardized independent variables vector for the i-th data point.

Next, CCA performs eigenvalue decomposition on the cross-covariance matrix C_{xy} to find its eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ and corresponding eigenvectors $(a_1, a_2, \ldots, a_k)$ [61]. CCA then sorts the eigenvalues in descending order and selects the top l eigenvectors associated with the l largest eigenvalues [62]. These eigenvectors represent the canonical correlation components [63]. CCA computes the canonical variables for both X and Y using the selected l eigenvectors [64]. For the i-th data point, the canonical variables u_i and v_i are calculated as Eqs. (28,29).

$$u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{il} \end{bmatrix} = x_i \cdot A_l \tag{28}$$

$$v_i = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{il} \end{bmatrix} = y_i \cdot B_l \tag{29}$$

where u_i and v_i are l-dimensional vectors representing the i-th data point in the reduced space for X and Y, respectively; x_i is the isotopes data point, y_i is the independent variables data point; and A_l and B_l are the matrices containing the selected l eigenvectors.

RESULTS AND DISCUSSION

Time series graphs of I-131, C-s134 and C-s137 in several locations

In this research, the author analyzed data from the 'DE.xlsx' file to track the evolution of radioactive isotopes (I-131, Cs-134, and Cs-137) formed during nuclear reactions. The data underwent multiple processing stages, including handling missing values and converting columns. Time series plots were then created to reveal temporal trends and changes in radioactivity levels, offering valuable insights into these isotopes' behavior over time. Time series graphs of I-131, Cs-134 and Cs-137 in several locations are shown in Fig. 10.

The research findings provide valuable insights into the historical trends of radioactive isotopes, enhancing our understanding of their patterns, characteristics, and potential environmental and human health impacts. Data processing techniques ensure result accuracy and reliability, while the graph displays isotope level trends at different locations over time, likely related to monitoring radiation effects from nuclear incidents or tests.

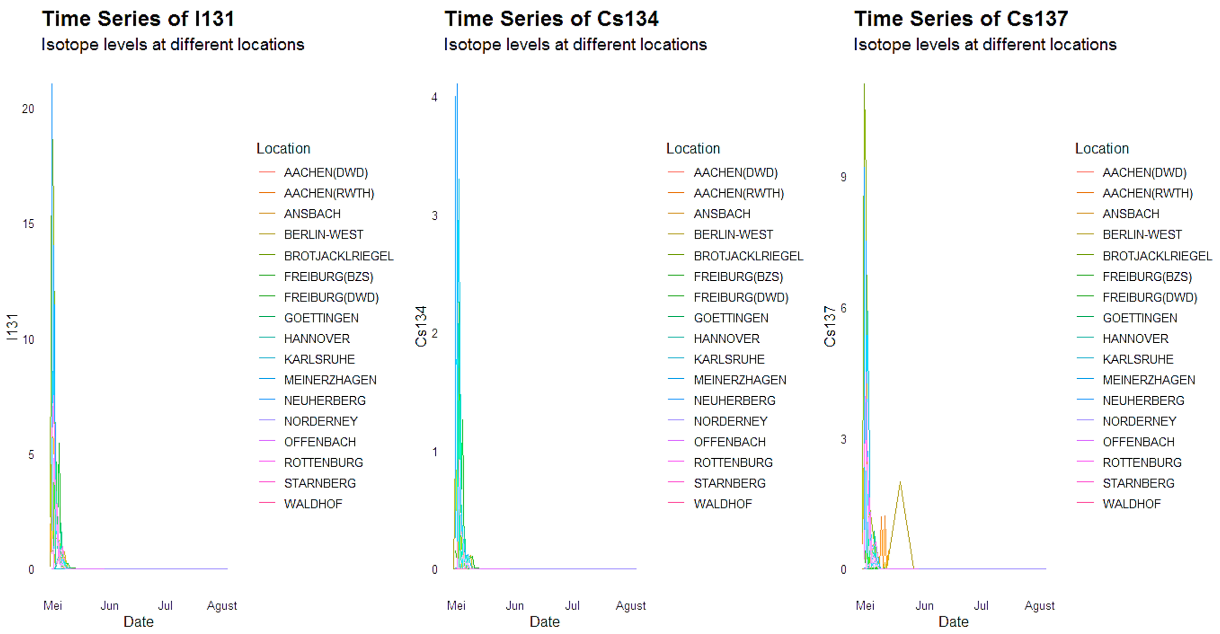


Fig. 10. Time series graphs of I-131, Cs-134 and Cs-137 in several locations.

Regression analysis of each isotope

This analysis examines the relationship between the dependent variable I-131 and the independent variables Longitude and Latitude. The linear regression model used in the analysis is as Eq. (30).

$$I131 = 18.94480 + 0.32541 \cdot \text{Longitude} - 0.41650 \cdot \text{Latitude} \quad (30)$$

The intercept of the model is 18.94480, and it is statistically significant with a p-value of 5.69×10^{-5}. The coefficient for Longitude is 0.32541, suggesting that for each unit increase in Longitude, the I-131 value increases by approximately 0.32541 units. This coefficient is also statistically significant with a p-value of 2.90×10^{-5}. The coefficient for Latitude is -0.41650, indicating that for each unit increase in Latitude, the I-131 value decreases by approximately 0.41650 units. This coefficient is statistically significant with a p-value of 7.30×10^{-6}.

The goodness of fit of the model is assessed using the Multiple R-squared value, which is 0.1402. It means that approximately 14.02 % of the variability in the I-131 values can be explained by the linear relationship with Longitude and Latitude. The Adjusted R-squared value is slightly lower at 0.1331, considering the degrees of freedom and the number of variables in the model. The F-statistic (19.74) with its associated p-value (1.148×10^{-8}) tests the overall significance of the model. Since the p-value is very low, we can conclude that the model as a whole is statistically significant in explaining the variation in I-131.

Next, this analysis examines the relationship between the dependent variable Cs-134 and the independent variables Longitude and Latitude. The linear regression model used in the analysis is as Eq. (31).

$$Cs134 = 4.455267 + 0.006229 \cdot \text{Longitude} - 0.085500 \cdot \text{Latitude} \quad (31)$$

The intercept of the model is 4.455267, and it is statistically significant with a p-value of 4.04×10^{-5}. The coefficient for Longitude is 0.006229, indicating that for each unit increase in Longitude, the Cs-134 value increases by approximately 0.006229 units. However, this coefficient is not statistically significant, as the p-value is relatively high at 0.724. The coefficient for Latitude is -0.085500, suggesting that for each unit increase in Latitude, the Cs-134 value decreases by approximately 0.085500 units. This coefficient is statistically significant with a p-value of 6.09×10^{-5}.

The goodness of fit of the model is assessed using the Multiple R-squared value, which is 0.06513. It means that approximately 6.513 % of the variability in the Cs-134 values can be explained by the linear relationship with Longitude and Latitude. The Adjusted R-squared value is slightly lower at 0.05737. The F-statistic (8.395) with its associated p-value (0.0002988) tests the overall significance of the model. The p-value is relatively low, indicating that the model as a whole is statistically significant in explaining the variation in Cs-134.

Finally, this analysis examines the relationship between the dependent variable Cs-137 and the independent variables Longitude and Latitude. The linear regression model used in the analysis is as Eq. (32).

$$Cs137 = 9.10266 + 0.17575 \cdot \text{Longitude} - 0.20289 \cdot \text{Latitude} \quad (32)$$

The intercept of the model is 9.10266, and it is statistically significant with a p-value of 0.000264. The coefficient for Longitude is 0.17575, suggesting that for each unit increase in Longitude, the Cs-137 value increases by approximately 0.17575 units. This coefficient is statistically significant with a p-value of 2.20×10^{-5}. The coefficient for Latitude is -0.20289, indicating that for each unit increase in Latitude, the Cs-137 value decreases by approximately 0.20289 units. This coefficient is statistically significant with a p-value of 3.77×10^{-5}.

The goodness of fit of the model is assessed using the Multiple R-squared value, which is 0.1315. It means that approximately 13.15 % of the variability in the Cs-137 values can be explained by the linear relationship with Longitude and Latitude. The Adjusted R-squared value is slightly lower at 0.1243. The F-statistic (18.32) with its associated p-value (3.91×10^{-8}) tests the overall significance of the model. The p-value is very low, indicating that the model as a whole is statistically significant in explaining the variation in Cs-137.

Descriptive analysis

This research investigated the relationships between three radioactive isotopes (I-131, Cs-134, and Cs-137) and various independent variables. Descriptive analysis revealed significant variability in the data for each isotope. For I-131, the average value was approximately 1.03 with a standard deviation of 2.72, while Cs-134 had an average value of approximately 0.22 with a standard deviation of 0.60. Cs-137 showed an average value of approximately 0.54 with a standard deviation of 1.44. These findings provide valuable insights into the characteristics of these isotopes. This condition is shown in Fig. 11.

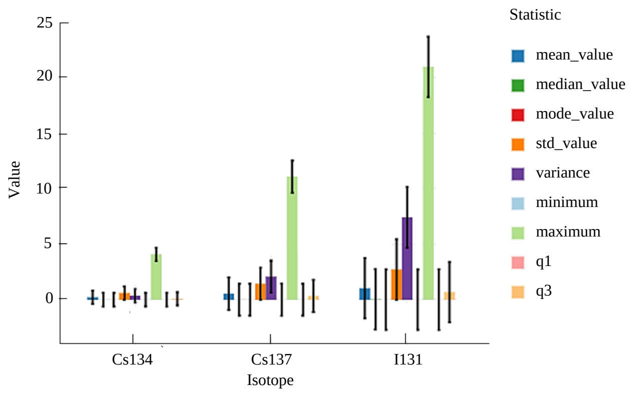


Fig. 11. Descriptive analysis bar diagram (mean, median, mode, standard deviation, variance, minimum, maximum, Q1, and Q3).

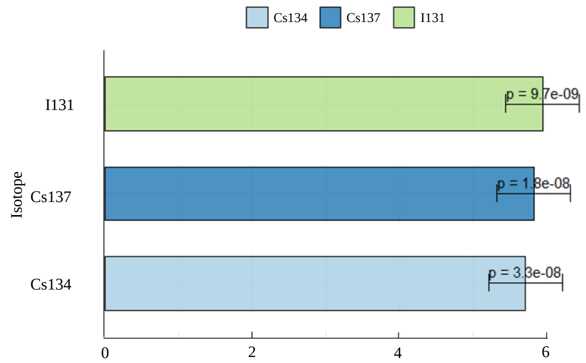


Fig. 12. The T-Values test results for sample means with confidence intervals.

One sample t-test

The research aimed to test hypotheses regarding three isotopes: I-131, Cs-134, and Cs-137. The method employed for this study was the One Sample t-test. The following are the research results for each isotope.

As shown in Table 2 and Fig. 12, the sample mean (mean of x) for the I-131 isotope was 1.033983. The t-statistic value was 5.9408, with degrees of freedom (df) equal to 244. The p-value obtained from the test was 9.723e-09 (extremely small). The 95 % confidence interval for the true mean of the I-131 isotope was between 0.6911542 and 1.3768123.

The sample mean (mean of x) for the Cs-134 isotope was 0.2195755. The t-statistic value was 5.7096, with degrees of freedom (df) equal to 243. The p-value obtained from the test was 3.29e-08 (extremely small). The 95 % confidence interval for the true mean of the Cs-134 isotope was between 0.1438232 and 0.2953278.

The sample mean (mean of x) for the Cs-137 isotope was 0.5363785. The t-statistic value was 5.8241, with degrees of freedom (df) equal to 244. The p-value obtained from the test was 1.803e-08 (extremely small). The 95 % confidence interval for the true mean of the Cs-137 isotope was between 0.3549726 and 0.7177844.

The results of the research indicated that the sample means for all three isotopes (I-131, Cs-134, and Cs-137) were significantly different from zero. The extremely small p-values obtained suggest strong evidence against the null hypothesis, indicating that the means of these isotopes are significantly different from the hypothesized values. The narrow 95 % confidence intervals also provide a precise estimate of the true means of each isotope.

Table 2. Statistical analysis of isotope data: t-values, degrees of freedom, p-values, 95 % confidence intervals, and sample means.

Isotope	t-value	Degrees of Freedom	p-value	95 % Confidence Interval	Sample Mean
I131	5.9408	244	9.72E-09	(0.6911542, 1.3768123)	1.033983
Cs134	5.7096	243	3.29E-08	(0.1438232, 0.2953278)	0.219576
Cs137	5.8241	244	1.80E-08	(0.3549726, 0.7177844)	0.536379

Based on the test results, it can be concluded that the p-value is very small for the three isotopes, namely I-131, Cs-134, and Cs-137. Therefore, the null hypothesis stating that the true mean of these three isotopes is 0 can be rejected. This means there is strong evidence to state that the true mean of these isotopes is not 0 and has statistically significant values. The 95 % confidence interval also provides an estimated range of values for the true mean of each isotope.

Pearson correlation coefficients

The Pearson correlation coefficient quantifies linear relationships between variables in Fig. 13, ranging from -1 (perfect negative) to 1 (perfect positive). I-131 and Cs-134 have a moderate positive correlation (0.56), I-131 and Cs-137 show a strong positive correlation (0.85), and Cs-134 and Cs-137 exhibit a moderate positive correlation (0.57). Spearman correlation, measuring monotonic associations, also ranges from -1 to 1. For I-131 and Cs-134, it's a very weak positive correlation (0.10); I-131 and Cs-137 display a moderate positive correlation (0.50); and Cs-134 and Cs-137 have a very weak to no correlation (-0.02) based on ranked data.

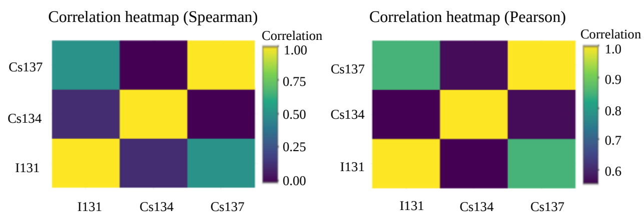


Fig. 13. Heatmap correlation using spearman and person methods.

Chi-square test

In this research, we conducted a Chi-Square test for three isotopes: I-131, Cs-134, and Cs-137.

As shown in Fig. 14, the Chi-Square test was performed to evaluate whether there is a relationship or association between the observed variables in the data. For the I-131 isotope, the Chi-Square test resulted in an X-squared value of 3074.3 with degrees of freedom (df) equal to 2544, and a p-value of 1.453e-12. The extremely small p-value indicates a significant relationship between the observed variables in this data. On the other hand, for the Cs134 isotope, the Chi-Square test yielded an X-squared value of 1633.3 with df equal to 1824, and a p-value of 0.9995. In this case, the p-value approaching 1 indicates that there is no significant relationship between the observed variables in the data. Moving on to the Cs-137 isotope, the Chi-Square test resulted in an X-squared value of 2625.5 with df equal to 2256, and a p-value of 8.173e-08. Similar to the I-131 isotope, the very small p-value suggests a significant relationship between the observed variables in this data.

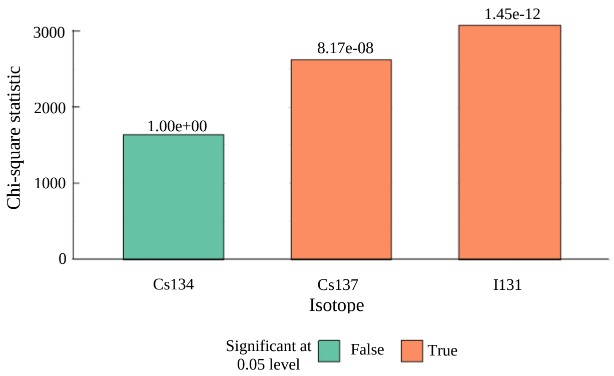


Fig. 14. Chi-square test results for different isotopes (pearson's chi-squared test).

Factor analysis

In Fig. 15, the study explored the connections between two factors, MR1 and MR2, and three variables, I-131, Cs-134, and Cs-137. It revealed the strength of these associations through standardized loadings: 0.71 for I-131, 0.57 for Cs-134, and 0.69 for Cs-137 in MR1, and 0.42 for I-131, 0.45 for Cs-134, and 0.62 for Cs-137 in MR2. The h2 values indicated the proportion of variance explained by the variables, with MR1 at 0.84 and MR2 at 0.38. Cs-137 had a strong relationship with its factor, with an h2 value of 0.86. Additionally, unique variances (u2) were determined for MR1 (0.16) and MR2 (0.62), with Cs137 having a u2 value of 0.14. Common variances (com) were calculated as 1.9 for MR1 and 2.0 for both MR2 and Cs-137. This analysis provides valuable insights into the data's underlying structure and associations.

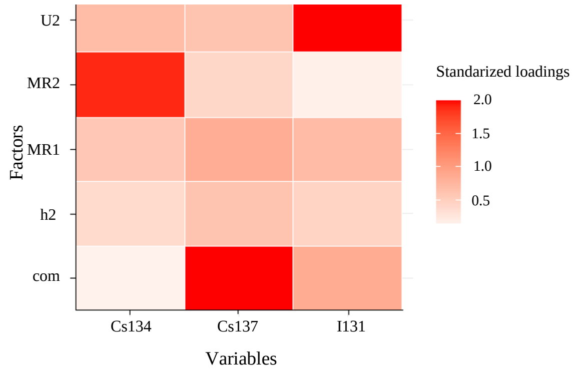


Fig. 15. Standardized loadings (pattern matrix).

The Fig. 16, presents the results of a Factor Analysis using the "minres" method with two factors, MR1 and MR2. It highlights factor loadings, variance explained, and cumulative variance. MR1 has factor loadings from 0.39 to 1.16, explaining 39 % of the variance, while MR2 has loadings from 0.31 to 0.92, explaining 31 % of the variance. When combined, they collectively explain 69 % of the variance. This analysis reveals crucial insights into the dataset's underlying structure and variable relationships.

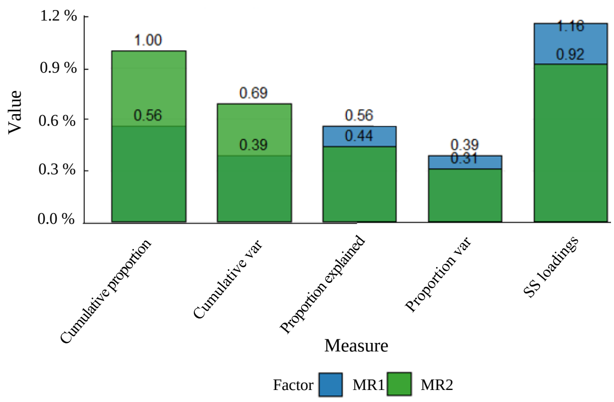


Fig. 16. Factor loadings, proportion var, and cumulative var.

As shown in Fig. 17, in effective research, factor analysis typically starts by establishing a null model, assuming no underlying factors in the data. Researchers then compare this null model with a 2-factor model to assess if adding factors significantly enhances model fit. The null model, with 3 degrees of freedom, has an objective function value of 1.68 and a chi-square value of 409.39. In contrast, the 2-factor model, which is overidentified, aims to minimize the objective function to zero for a perfect fit. Interpreting these results carefully is crucial, as a perfect fit like the one observed here is uncommon in practice. Researchers may also consider other fit indices for a comprehensive model assessment.

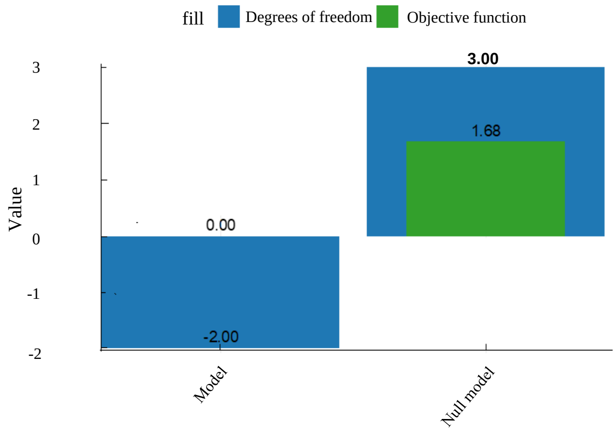


Fig. 17. Minres method indicate that two main factors can be identified from the data used.

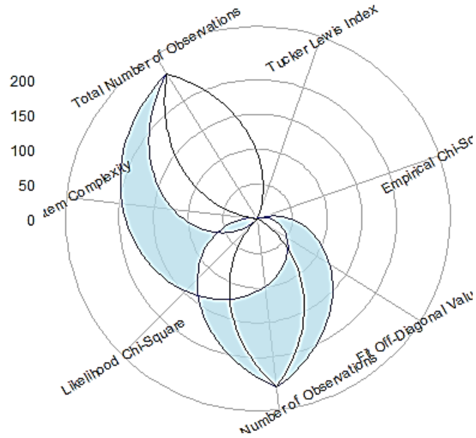


Fig. 19. Aditional information radar plot: values for differents metrics.

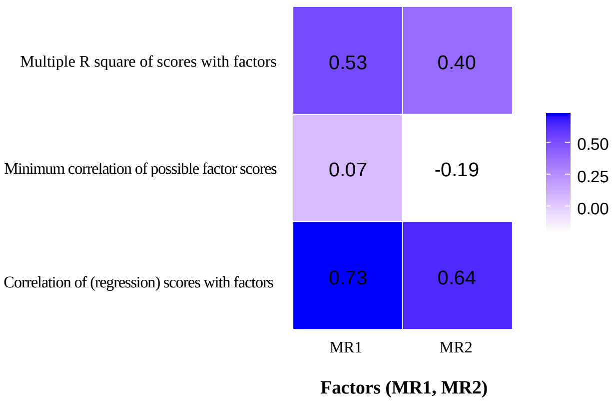


Fig. 18. Measures of factor adequacy.

As shown in Fig. 18, the research findings using factor analysis with the minres method reveal two main factors, labeled as MR1 and MR2. MR1 shows a strong correlation of 0.73 with regression scores, explaining about 53 % of the score variation. MR1 also has minimal correlations with other factors, making it conceptually distinct. On the other hand, MR2, with a correlation of 0.64, explains around 40 % of score variation but is closely related to other factors, as indicated by its negative correlation. These factors offer valuable insights into the data relationships, but interpreting them should consider the context and research objectives.

The pattern matrix from the correlation matrix revealed significant associations between variables and factors. Factor MR1 exhibited strong connections with I-131 (0.71) and Cs-137 (0.69), and a moderate one with Cs-134 (0.42). Factor MR2 displayed substantial links with Cs-134 (0.45) and Cs-137 (0.62), and a somewhat lower one with I-131 (0.57).

The h2 values show that factors explain a significant portion of variance in observed variables. Factor MR1 accounts for 84 % of the variance, while Factor MR2 explains 86 %. Low u2 values indicate minimal unexplained variance, suggesting a good model fit. Communalities for variables are 1.9 (I-131) and 2.0 (Cs-134 and Cs-137). The table provides vital data on factor loadings, variance, and cumulative variance. MR1 has a sum of squares loadings of 1.16, explaining 39 % of the variance, while MR2 accounts for 31 %. Together, they explain 69 % of the variance. Statistical tests confirm the model's goodness-of-fit, with perfect fit indicators and a strong factor relationship. The factors are distinct and reliable, explaining 53 % (MR1) and 40 % (MR2) of the variance in scores. Low correlations between factor scores support their distinctiveness (0.07 for MR1, -0.19 for MR2). This condition is shown in Fig. 19.

Principal component analysis (PCA) and canonical correlation analysis (CCA)

This research focused on analyzing the Eigenvalues of three dimensions (Dim.1, Dim.2, and Dim.3) within a specific context. The findings revealed that Dim.1's Eigenvalue of 2.323, with a variance percentage of 77.440 %, plays a significant role in representing the data. Dim.2, with an Eigenvalue of 0.524 (17.473 % of total variance), also contributes to data variability. Dim.3, though smaller, holds significance with an Eigenvalue of 0.153 (5.088 % of total variance). Cumulatively, Dim.1 and Dim.2 encompass 94.912 % of data variability, while Dim.3 completes the cumulative variance to 100.000 %, effectively capturing the entire data variability in this research.

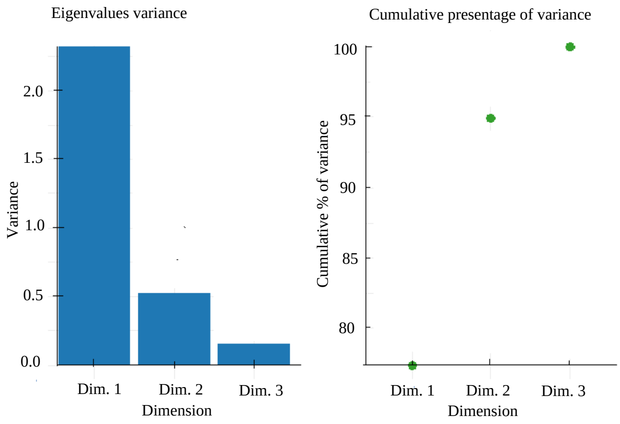


Fig. 20. Eigenvalues variance and cumulative percentage of varianced.

As shown in Fig. 20, this research employed Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) to explore the connection between isotopes and geographical factors (Longitude and Latitude). In Fig. 21, PCA revealed two main components, Dim.1 and Dim.2, explaining 48.8 % of isotopic variation. Dim.1 clarified 28.8 % of the variation with a 1.5 % ctr and 13.1 % cos2. Dim.2 explained 23.1 % of the variation with a 2.4 % ctr and 18.4 % cos2. Dim.3, though generated by PCA, only explained 6.7 % with a 0.5 % ctr and 39.6 % cos2. The cumulative variance explained by Dim.1 and Dim.2 is around 48.8 %. CCA examined the relationship between isotopes and Longitude/Latitude, involving 10 individuals or data points. Dist represents the distance to the CCA center. Dim.1, Dim.2, and Dim.3 are CCA components. For each individual, CCA values and component variables are provided. CCA value indicates the association with each component, ctr reveals variation explanation, and cos2 shows the proportion of variance explained.

This research utilized multivariate analyses, including Principal Component Analysis (PCA) and Canonical Correspondence Analysis (CCA), to investigate the associations between radioactive isotopes and geographic variables like Longitude and Latitude. In Fig. 22, PCA Dimension 1 (Dim.1) explained 92.3 % of the variability in isotope I-131, with a strong correlation (cos2 = 0.851). Dim.2 and Dim.3 had weaker relationships (cos2 = 0.074 and 0.075, respectively). In CCA, Dim.1 contributed 36.6 % and had a strong correlation (ctr = 0.923) with isotope I-131. Dim.2 and Dim.3 had lower relationships (ctr = -0.271 and 0.274). PCA analysis on Cs-134 and Cs-137 showed weaker correlations in the second and third principal components.

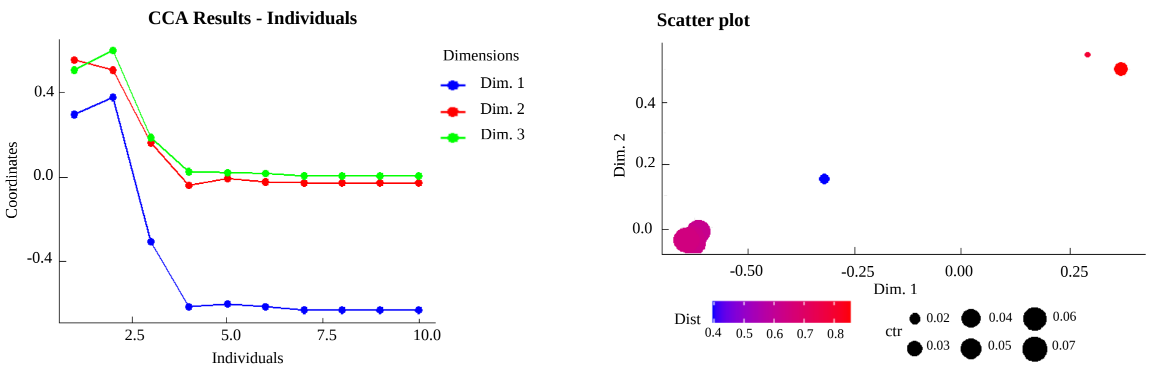


Fig. 21. CCA results graph.

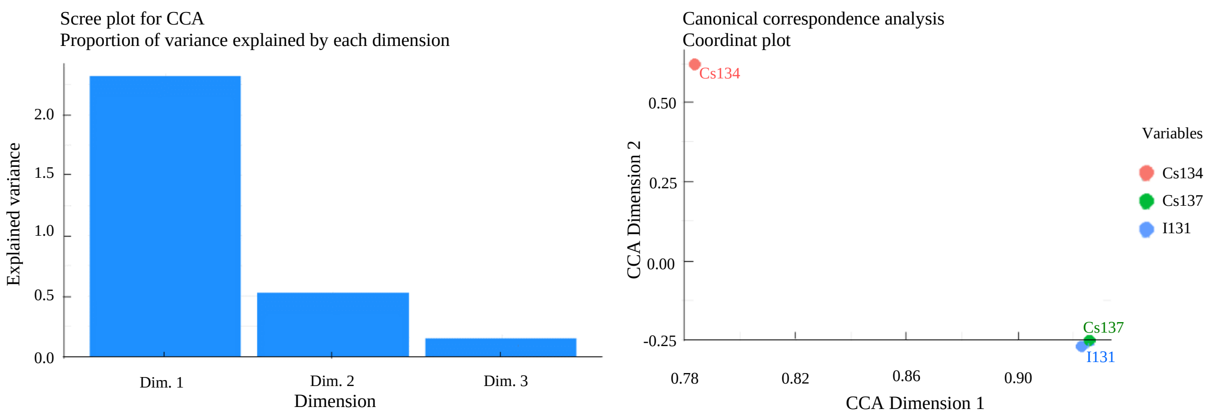


Fig. 22. Scree plot for CCA for proportion of variance explained by each dimension and canonical correspondence analysis in coordinat plot.

In this study, a comprehensive analysis of the behavior and relationships of three radioactive isotopes (I-131, Cs-134, and Cs-137) over time was conducted. Data from the "DE.xlsx" file was utilized to depict the progression of these isotopes. Time series graphs revealed trends and temporal variations in radioactivity levels at various locations. These visualizations offer valuable historical insights into the behavior of these radioactive isotopes, with potential implications for the environment and human health.

Regression analysis investigated the relationship between radioactive isotope levels (dependent variable) and geographic coordinates, Longitude, and Latitude (independent variables). The results revealed statistically significant associations between Longitude, Latitude, and the levels of I-131 and Cs-137, indicating a geographic factor in their distribution. However, Cs-134 did not show a statistically significant relationship with Longitude, suggesting unique characteristics.

Descriptive analysis provided important statistical metrics for all three isotopes, including mean, median, mode, standard deviation, variance, minimum and maximum values, and quartiles. These statistics offer a comprehensive view of the distribution and variability of these isotopes.

One-sample t-tests aimed to assess whether the mean isotopic values significantly differed from zero. Very low p-values indicated strong evidence against the null hypothesis, confirming that the isotopes have statistically significant mean values. Precise confidence intervals provided accurate estimates of the true means for each isotope.

Pearson and Spearman correlation coefficients revealed varying levels of correlation between isotopes, indicating diverse degrees of association. This information sheds light on the interactions between these isotopes.

Chi-Square tests evaluated relationships among observed variables (isotopes) in the dataset. Significant relationships were found for I-131 and Cs-137, indicating their interrelated behavior. However, Cs-134 did not show a significant relationship, signifying its unique behavior.

Factor analysis aimed to identify underlying factors (MR1 and MR2) and their relationships with the isotopes. Standardized loadings, h2 values, u2 values, and common variance provided insights into the strength and relationships of these factors. The analysis revealed two primary factors (MR1 and MR2) explaining a significant portion of the variance in isotopes.

Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) were employed to assess the relationships between isotopes and geographic variables (Longitude and Latitude). Eigenvalues and variance percentages highlighted the importance of each dimension in explaining data variability, while CCA revealed the extent to which isotopes were associated with independent variables.

This research extends previous studies by offering a deeper and more comprehensive analysis of the behavior and relationships of these radioactive isotopes. It provides new insights into their associations with geographic factors, correlations between isotopes, and underlying factors influencing their behavior. These findings contribute to a deeper understanding of isotopic behavior and its implications, making it a valuable addition to the existing knowledge in this field.

CONCLUSION

In conclusion, this study provides a comprehensive analysis of changing radioactivity levels of isotopes (I-131, Cs-134, and Cs-137) at various locations over time. The research revealed significant associations between I-131 and Cs-137 levels with Longitude and Latitude, emphasizing the role of environmental factors in their distribution. Descriptive statistics, one-sample t-tests, Pearson and Spearman correlation coefficients, and Chi-Square tests all confirmed the statistical significance and relationships among the isotopes. Factor analysis identified two key factors (MR1 and MR2) driving isotopic variations, shedding light on the forces shaping radioactivity distribution. The integration of PCA and CCA provides a comprehensive approach to understanding isotopes in relation to independent variables, offering valuable insights for future research.

Furthermore, the study underscores the importance of establishing a robust environmental monitoring system to track radioactive isotopes in various locations, enabling timely detection of changes in radioactivity levels. This is particularly crucial in areas with higher isotopic concentrations, as it allows for health impact assessments and effective risk communication strategies to inform the public about potential risks. Further research should delve into the specific impacts of geographic features on individual isotopes, given their varying correlations, and explore the nature and significance of the factors influencing isotopic variations. These findings lay the groundwork for a deeper understanding of the complex interactions involved in radioactivity distribution and environmental monitoring.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the anonymous reviewers for their insightful comments and valuable suggestions that significantly contributed to the enhancement of this research paper. Special appreciation is extended to the institutions and departments affiliated with the authors for their support and facilitation throughout the study. The collaborative efforts of the research team from Universitas Negeri Medan, University of Mataram, Sam Ratulangi University, and Politeknik Kelautan dan Perikanan Pangandaran are acknowledged with deep appreciation. This research would not have been possible without the collective commitment and dedication of all involved parties.

AUTHOR CONTRIBUTION

Budiman Nasution conceptualized and designed the research, formulated the analytical framework, and played a crucial role in data interpretation. Winsyahputra Ritonga conducted data acquisition, processed and analyzed time series data, and contributed to the drafting of the manuscript. Aris Doyan contributed to the methodological design, performed statistical analyses using Linear Regression, and assisted in interpreting the results. Paulus Dolfie Pandara played a significant role in data preprocessing, imputation of missing values, and contributed to the critical revision of the manuscript. Lulut Alfaris contributed to the interpretation of results from Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) and provided valuable insights into the geographical aspects of the study. Ruben Cornelius Siagian contributed to the literature review, manuscript drafting, and ensured the coherence of the research narrative. All authors critically reviewed and approved the final version of the manuscript, acknowledging their individual roles in this comprehensive analysis of temporal trends and spatial relationships of radioactive isotopes.

REFERENCES

1. I. Zuba, M. Zuba, M. Piotrowski *et al.*, Appl. Radiat. Isot. **162** (2020) 109176.
2. Q. T. Birch, P. M. Potter, P. X. Pinto *et al.*, Talanta **224** (2021) 121743.
3. L. Terzi, G. Wotawa, M. Schoeppner *et al.*, Sci. Rep. **10** (2020) 1.
4. M. Aldener, A. Axelsson, T. Fritioff *et al.*, J. Environ. Radioact. **262** (2023) 107159.
5. P. A. Assimakopoulos, K. G. Ioannides, Pakou *et al.*, J. Dairy Sci. **70** (1987) 1338.
6. S. Hashimoto, M. Komatsu and S. Miura, Forest Radioecology in Fukushima, Springer Nature Singapore, Singapore (2022) 1.
7. O. Masson, J. Tschiersch, L. S. Lebel *et al.*, *Radioiodine Releases in Nuclear Emergency Scenarios*, in: Nuclear Emergencies, Springer, Japan (2019) 175.
8. J. Iqbal, F. M. Howari, A. M. O. Mohamed *et al.*, *Assessment of Radiation Pollution from Nuclear Power Plants*, in: Pollution Assessment for Sustainable Practices in Applied Sciences and Engineering, Elsevier, United States (2021) 1027.
9. S. Kang, Q. Zhang, Y. Qian *et al.*, Natl. Sci. Rev. **6** (2019) 796.
10. C. Selvan and S. R. Balasundaram, *Data Analysis in Context-Based Statistical Modeling*, in: Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics, IGI Global, United States (2021) 96.
11. S. Senthilnathan, Usefulness of Correlation Analysis. https://papers.ssrn.com/sol3/papers.cfm? abstract_id=3416918. Retrieved in February (2023).
12. P. Schober, T. R. Vetter, Anesth. Analg. **132** (2021) 365.
13. I. J. B. D. Nascimento, D. P. O. Mathúna, T. C. V. Groote *et al.*, BMC Infect. Dis. **21** (2021) 525.
14. K. S. J. Glew, B. Espinasse, B. P. V. Hunt *et al.*, Trueman, Global Biogeochem. Cycles **35** (2021) 1.
15. P. Mishra, C. Pandey, U. Singh *et al.*, Ann. Card. Anaesth. **22** (2019) 297.
16. G. Francis, V. Jakicic, Behav. Res. Methods **55** (2023) 77.
17. A. Beyer, *Introduction to t-tests*, in: Introduction to Statistics for Psychology, Pressbooks, Canada (2021) 1.
18. P. Maneejuk and W. Yamaka, J. Appl. Stat. **48** (2021) 827.
19. C. Andrade, Indian J. Psychol. Med. **41** (2019) 203.
20. L. Hespanhol, C. S. Vallio, L. M. Costa *et al.*, Braz. J. Phys. Ther. **23** (2019) 290.
21. Z. Hu, Y. Yang, G. Li *et al.*, Scand. J. Stat. **50** (2023) 1884.

22. J. Salupeto-Dembo, Z. Szabó-Krausz, P. Völgyesi *et al*., J. Radioanal. Nucl. Chem. **325** (2020) 271.
23. H. S. Darling, Cancer Res. Stat. Treat. **5** (2022) 139.
24. R. A. Armstrong, Ophthalmic Physiol. Opt. **39** (2019) 316.
25. B. Pratap, S. Mondal and B. H. Rao, Asian J. Civ. Eng. (2023) 1.
26. E. Seeram, Radiol. Technol. **91** (2019) 176.
27. G. Perinetti, South Eur. J. Orthod. Dentofac. Res. **6** (2019).
28. M. J. Osei and J. Kim, Econ. Modell. **93** (2020) 154.
29. E. Temizhan, H. Mirtagioglu and M. Mendes, Am. Acad. Sci. Res. J. Eng. Technol. Sci. **85** (2022) 265.
30. J. Hou, X. Ye, W. Feng *et al*., BMC Bioinf. **23** (2022) 81.
31. A. Shekhar, S. Patel, Int. J. Res. Appl. Sci. Eng. Technol. **10** (2022) 653.
32. H. Mohajan, J. Econ. Dev. Environ. People **9** (2020) 50.
33. Q. Yang, A. M. Williamson, A. Hasted *et al*., Food Qual. Preference **83** (2020) 103928.
34. R. J. Smith, Am. J. Phys. Anthropol. **172** (2020) 521.
35. N. S. Turhan, Educ. Res. Rev. **15** (2020) 575.
36. M. Whatley, Introduction to Quantitative Analysis for International Educators, Springer, USA (2022) 57.
37. J. Shrestha, P-Value: a True Test of Significance in Agricultural Research https://www.linkedin.com/ pulse/p-value-test-significance-agricultural-rese arch-jiban-shrestha/. Retrieved in February (2023).
38. A. E. Kossovsky, Stats **4** (2021) 419.
39. N. Shrestha, Am. J. Appl. Math. Stat. **9** (2021) 4.
40. N. Shrestha, Am. J. Appl. Math. Stat. **8** (2020) 39.
41. A. Robitzsch, Front. Educ. **5** (2020) 1.
42. E. Bergqvist, T. Tossavainen and M. Johansson, Educ. Res. Int. **2020** (2020) 1.
43. B. G. Hansen and O. Østerås, Prev. Vet. Med. **170** (2019) 104741.
44. G. D. Garson, Factor Analysis and Dimension Reduction in R, Routledge, London (2022) 1.
45. J. M. D. Ampong, J. A. Bagares, J. A. Berja *et al*., Can. J. Fam. Youth **15** (2023) 90.
46. D. Almaleki, Eng. Technol. Appl. Sci. Res. **11** (2021) 6849.
47. M. J. Kochenderfer and T. A. Wheeler, Algorithms for Optimization, The MIT Press, London (2019) 1.
48. G. Pavlov, D. Shi and A. Maydeu-Olivares, Struct. Equation Model. **27** (2020) 908.
49. C. Lasanthika, K. Wanigasuriya, U. Hettiaratchi *et al*., J. Patient Exper. **10** (2023) 1.
50. A. Grant, S. Lagdon, J. Devaney *et al*., PLoS One **18** (2023) e0285835.
51. Y. Takane and H. Hwang, Multivar. Behav. Res. **37** (2002) 163.
52. A. Balasubramaniam, M. J. J. Gul, V. G. Menon *et al*., IETE Tech. Rev. **38** (2021) 438.
53. B. M. S. Hasan and A. M. Abdulazeez, J. Soft Comput. Data Min. **2** (2021) 20.
54. F. Kherif, A. Latypova, *Principal Component Analysis*, in: Machine Learning, Elsevier, United Kingdom (2020) 209.
55. J. F. Hair, M. C. Howard and C. Nitzl, J. Bus. Res. **109** (2020) 101.
56. N. Kettaneh, A. Berglund and S. Wold, Comput. Stat. Data Anal. **48** (2005) 69.
57. D. N. Lawley and A. E. Maxwell, J. R. Stat. Soc. D 1**2** (1962) 209.
58. N. Salem and S. Hussein, Procedia Comput. Sci. **163** (2019) 292.
59. Y. Li, J. Ren, Y. Yan *et al*., J. Phys. Conf. Ser. **2278** (2022) 012021.
60. K. M. M. Uddin, N. Biswas, S. T. Rikta *et al*., Comput. Methods Programs Biomed. Update **3** (2023) 100098.
61. J. W. Choi, F. Zotter, B. Jo *et al*., IEEE/ACM Trans. Audio Speech Lang. Process. **30** (2022) 2310.
62. K. Martinkus, A. Loukas, N. Perraudin *et al*., *SPECTRE: Spectral Conditioning Helps to Overcome the Expressivity Limits of One-shot Graph Generators*, Proceedings of the 39th International Conference on Machine Learning (2022) 15159.
63. X. Zhuang, Z. Yang and D. Cordes, Hum. Brain Mapp. **41** (2020) 3807.
64. W. Wang and Y. H. Zhou, J. Multivar. Anal. **185** (2021) 104781.