WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A REVIEW OF VARIOUS METHODS OF CORRELATIONS ANALYSIS ON WATER QUALITY PARAMETERS

Kapil Naresh Dwivedi[a], Aditya Lakhera[b]*

[a]Research Scholar, Shri Krishna University, Chhatarpur, 471001, India
[b]Assistant Professor,, Shri Krishna University, Chhatarpur, 471001, India

A B S T R A C T

In this paper, various methods of correlation analysis were discussed. The correlation analysis was conducted to assessed the quality and pollution status of source surface waters of Narmada Jabalpur, by monitoring the nature, cause, and extent of pollution in the Annuper Region. The data processed collected over a period of 12 months, between Jabalpur and December 2019 made available by CPCB and again for a month in 2021. A total of 11 kinds of physicochemical parameters were assessed. The objective is to determine significantly important parameters by statistical analysis using the Karl-Pearson method. We were able to conclude that Major Correlations in the Narmada river were Turbidity & Total Solids (0.815); Total hardness & Sp. Conductivity (0.847); BOD & Sp. Conductivity (0.831); COD & Nitrate (0.819); Total Hardness & DO (0.883). A review of various methods available for correlation analysis is done in this paper.

Keywords*: Correlation Analysis, Water Quality Index, Narmada River, Karl-Pearson Method, Fuzzy Logic*

1. INTRODUCTION

Correlation analysis is one of the most widely used and reported statistical methods in summarizing the last amount of research data. The Correlation coefficient or r coefficient is a statistical term used to measure the degree or strength of the correlation between two variables. It is also referred to as Pearson's product-moment r or r coefficient. The value can range in both magnitude and direction being either positive or negative. It may take a range of values from -1 to 0 to +1. The strength of the correlation coefficient is not dependent on the direction of the sign. The R-value of +0.80 is equally correlated to the -0.80 value. It is so that a positive correlation coefficient indicates that an increase in the first variable would correspond to an increase in the second variable. Negatives mean the inverse relationship. An example of a negative correlation might involve a person doing the pull-ups relative to the percentage of body fat, meaning if pull-ups increase body fat decreases and if the pull-up increases. The principle of correlation analysis is used to find out the most correlated/important physicochemical parameter so that it can be carefully monitored and other parameters would also be in a specified range finally resulting in a much cleaner river/water body.

An illustration displayed below shows, that the correlation remains higher as the data observations fall closer to a straight line and the coefficient value decreases as the data point deviates more from the straight line. No correlation would result in random scattering of the data points. In statistical analysis, it is very important to properly interpret the correlation. Generally, r values (in absolute values) r<= 0.35 is considered a low or weak correlation. 0.36 to 0.37 is modest and 0.68 to 1.0 is considered a strong or high correlation with r > 0.90 as a very high correlation. The correlation coefficient indicates how closely the data fit in a linear pattern as shown in figure(….). Correlation analysis goes further by determining the existing relationship between the parameters known as *Regression analysis*. In Regression analysis, a mathematical equation is developed for the line representing the best fit around the data. And from this equation predictions can be made. Unknown values or dependent variables are predicted from independent variables or known values.

2. REVIEW OF RELATED STUDIES

2.1 Correlation Analysis of Leachate in final disposal sites on groundwater and surface water quality

The goal of this research was to determine the level of groundwater and surface water pollution as well as the correlation of leaches to groundwater and surface water in west Lombok's final disposal site in Suka Makmur village Gerung district. Pollution index and Pearson Correlation were used and the results indicated the adequate condition of groundwater wells and TSS correlation test analysis on groundwater turbidity came out to be a strong correlation of 0.643. With Surface-water it was 0.357. COD & BOD of Well water against Groundwater came out as 0.257 & 0.475.

2.2 Evaluating the surface water quality index fuzzy and its influence on water treatment

In this paper, Fuzzy based correlation analysis was performed and a new Raw water quality index was developed in compliance with deterministic models, to assess the treatability of natural waters for human consumption using the conventional process, this is named Raw Water Quality Index Fuzzy or RWQIF. The dataset of 24 water sources in Brazil was used. This model proves to be a suitable decision-making tool that allows managers to prioritize watershed protection actions aiming lower water treatment costs and improve treated water quality.

2.3 Correlation Study among Water Quality Parameters of Groundwater of Valsad District of South Gujarat (India).

In this study, Groundwater samples from 5 talukas of the Valsad district from August 2008 to July 2009 were analyzed using Karl Pearson Method. The parameters were pH, Colour, Electrical Conductivity, Total Hardness, Total Dissolved Solids, Silica, Chloride, Sulphate, Fluoride, Sodium, Chemical oxygen demand, and metals like Copper and Manganese. Using correlation coefficients highly correlated parameters and interrelated water parameters were determined. The study determined that EC is highly correlated with 8 out of the other 17 water quality parameters.

2.4 Correlation Study for the Assessment of Water Quality and its Parameter of Ganga River, Kanpur, Uttar Pradesh, (India).

This study was very similar to a study conducted in the Valsad district. This research was conducted on the Ganga river in Kanpur at 6 different locations from March 2010 to February 2011. The data were collected monthly, water quality assessment was carried out and highly correlated parameters were determined by Pearson's Correlation Coefficient value (r). They also determined the p-value for various pairs to test the joint effect of several independent variables. Their results showed that Chromium was non-significant and could be avoided in Physico-chemical analysis of water. Their results also displayed that mean values of all physicochemical parameters were within the highest desirable limit set by World Health Organization.

2.5 Correlation Study on Physio-chemical Parameters and Quality Assessment of Kosi River Water, Uttarakhand.

This study was done on the Kosi river at the Kosi sampling station during pre-monsoon and post-monsoon in the year 2004-2005. Statistical studies were carried out by calculating correlation coefficients between different pairs and a t-test was applied for checking significance. The research concluded that parameters were within the recommended WHO limits. The paper also concluded that Chloride has a positive correlation with pH, Mg, Na, Hardness, and total suspended solids. A negative correlation was discovered between Sodium with Hardness, EC, and Turbidity.

3. REVIEW OF VARIOUS METHODS OF CORRELATION ANALYSIS

3.1 Regression Analysis

Regression Analysis is the oldest and most widely used multivariate technique in statistics. In regression analysis, the objective is to obtain a prediction of one variable by giving the value of other variables. This method usually helps us in the following:

- When we want to know if there is any relationship between two or more variables, actually exists or not.
- When we want to understand what kind of relationship exists.
- When we want to predict a variable by giving the value of other variables.

The simplest form of Regression is very similar to correlation analysis, but in regression, there is another added step of figuring out the actual Mathematical equation which would tell the current relationship between the variables.

The word "Regression" was first used by Francis Galton in the 19th century for describing a biological phenomenon of the height of descendants relative to their tall ancestors.

There is Linear Regression or simple linear regression, General Linear Regression, Non-linear Regression, Nonparametric Regression & Quantile Regression, etc.

3.2 Karl Pearson's Coefficient of Regression

It was first introduced by Karl Pearson from a related idea earlier give by Francis Galton in the year 1880. Karl Pearson's coefficient of correlation is used in mathematical methods and statistics to find the level of relation between linearly related variables. The coefficient is denoted by 'r'. There are multiple variants of this formula but that is out of scope for this study.

The coefficient can be calculated by the following formula:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where 'n' is the sample size, 'xi' 'yi', are the variables, $\bar{x}$, $\bar{y}$ are mean of x and y.

The value of the Karl Pearson coefficient can range from -1 to 0 to +1. The following points explain their significance:

- When the correlation coefficient is +1, it means that for every increase in the independent variable, there will be a positive increase in the dependent variable in a fixed proportion. For example, shoe size increases as the length of feet increases.
- When the correlation coefficient is -1, it means that for every increase in the independent variable, there will be a negative decrease in the dependent variable in a fixed proportion. For example, the amount of gas in a kitchen cylinder decreases as we start cooking something on the stove.
- When the correlation coefficient is 0, it means for every increase in the independent variable, there is no positive or negative change independent variable. It can also be said that the variables are not correlated.

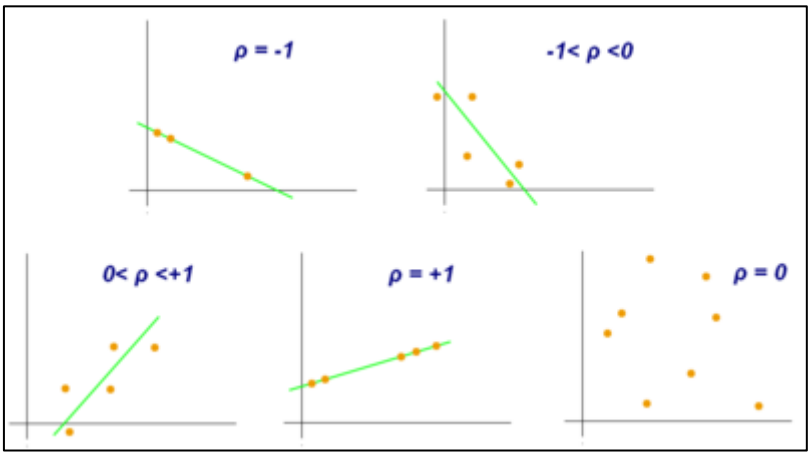

Figure 1 : Examples of scatter diagrams with different values of the correlation coefficient.

3.3 Spearman Correlation Coefficient

The Spearman's rank coefficient of correlation is a measure of nonparametric rank correlation or dependence of ranking between two variables. It was introduced by Charles Spearman a psychologist from the UK in 1904. It is very similar to Karl Pearson's coefficient but the Pearson coefficient works with a linear relationship between the two variables whereas the Spearman coefficient works with a monotonic relationship between the variables.

A monotonic relationship is when:

- As one variable increases, so does the other,
- As one variable increases, the other variable decreases, But NOT exactly at a constant rate wherein linear relationship the rate of increase or decrease is very much constant.

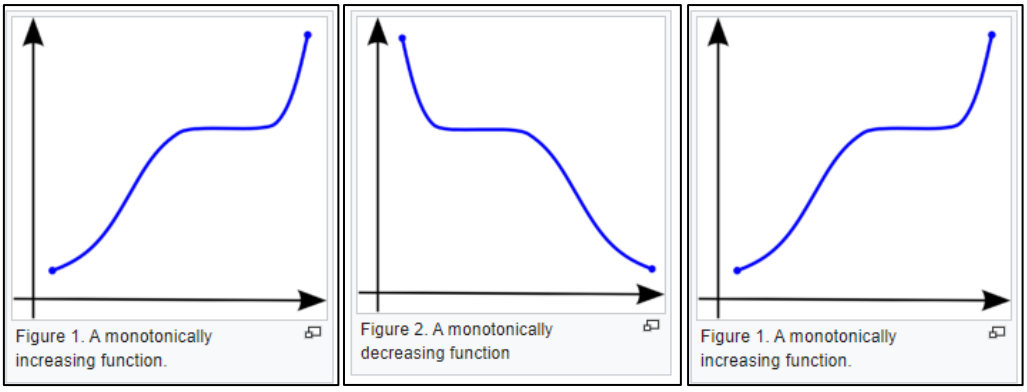

Figure 2 : Examples of Monotonic relationships between the variables.

The Spearman coefficient can be calculated by the following formula:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Rho – Spearman's Rank Correlation Coefficient

Di = difference between the ranks

N = number of observations.

3.4 Fuzzy Logic

The application of Fuzzy based correlation analysis with fuzzy set theory has always been a measure of uncertainty either qualitative or approximate as the research area for the analysis. Fuzzy-based correlation is reconsidered from correlation coefficient into fuzzy settings employing the observation values or either by qualitative understanding in general terms such as 'bad', 'good', or excellent. The calculations in fuzzy correlations are employed by two different methods, the first which relies on Zadeh's extension principle finding membership function. Membership function can be found by various methods for example using t-norms and nonlinear programming. Fuzzy helps comprehend the relation between two variables that are correlated either in general terms or in approximate quantities.

REFERENCES

[1] De Side, G. N., Widiyanti, A., Rancak, G. T., Aprianto, R., Widhiantari, I. A., & Sutawijaya, I. B. (2021, November). Correlation analysis of leachate in final disposal sites on groundwater and surface water quality. In IOP Conference Series: Earth and Environmental Science (Vol. 913, No. 1, p. 012048). IOP Publishing.

[2] de Oliveira, M. D., de Rezende, O. L. T., de Fonseca, J. F. R., & Libanio, M. (2019). Evaluating the surface Water quality index fuzzy and its influence on water treatment. Journal of Water Process Engineering, 32, 100890.

[3] Shroff, P., Vashi, R. T., Champaneri, V. A., & Patel, K. K. (2015). Correlation study among water quality parameters of groundwater of Valsad district of south Gujarat (India). Journal of Fundamental and Applied Sciences, 7(3), 340-349.

[4] Khatoon, N., Khan, A. H., Rehman, M., & Pathak, V. (2013). Correlation study for the assessment of water quality and its parameters of Ganga River, Kanpur, Uttar Pradesh, India. IOSR Journal of Applied Chemistry, 5(3), 80-90.

[5] Bhandari, N. S., & Nayal, K. (2008). Correlation study on physico-chemical parameters and quality assessment of Kosi river water, Uttarakhand. E-Journal of Chemistry, 5(2), 342-346.