Environ Res Tec, Vol. 5, Issue. 2, pp. 165–171, June 2022

Environmental Research and Technology
https://ert.yildiz.edu.tr - https://dergipark.org.tr/tr/pub/ert
DOI: https://doi.org/10.35208/ert.1062683

Research Article

Principal component analysis in long term assessment of total viable plate count of municipal water distribution network system in healthcare facility

Mostafa ESSAM EISSA[*1], **Engy REFAAT RASHED**[2], **Dalia ESSAM EISSA**[3]

[1]*Independent Researcher, Pharmaceutical and Healthcare Research Facility, Cairo, Egypt*
[2]*National Centre for Radiation Research and Technology, Cairo, Egypt*
[3]*Royal Oldham Hospital, Oldham, United Kingdom*

ARTICLE INFO

Article history
Received: 25 January 2022
Revised: 19 April 2022
Accepted: 27 April 2022

Key words:
Eigen value; Loading plot; Outlier diagram; PCA; Score graph; Scree plot

ABSTRACT

The microbial quality of water is a critical safety aspect that should not be overlooked, especially when considering human consumption in the facilities for healthcare and the treatment of health compromised populations. The present work aimed to study a municipal network for city water within a healthcare facility microbiologically. The implementation of multivariate analysis was conducted over collected data to decipher trends of the microbiological count of samples from different points-of-use. The statistical study involved an Eigenvalue determination, score graph, loading plot study and outlier diagram. Data showed clustering tendency with aberrant values could be observed. The first component appeared to be associated with bioburden density in the water. While the other component showed a link to the relative locations of the distribution points in the facility and the length of the water lines from the source feeding city water to a great extent. The two components explained more than 55% of the variations in data. A property that highlighted a strong agreement between the order of points-of-use in the loading plot with that of ratio factor which was obtained from the Total Viable Count over the hypothetical distance of sampling port from the reference feeding entrance. Thus, there was evidence of variation in the microbial quality affected by the biological stability of the water distribution system that needed attention and an improvement plan for better control of microbial content in city water to prevent possible excursions in the future. The study showed a valuable perspective approach for the quantitative examination of the microbiological stability of the municipal water distribution network system.

Cite this article as: Essam Eissa M, Refaat Rashed E, Essam Eissa D. Principal component analysis in long term assessment of total viable plate count of municipal water distribution network system in healthcare facility. Environ Res Tec 2022;5:2:165–171.

INTRODUCTION

The microbiological quality of water is a crucial criterion that must be met for safe human use and consumption of city water as could be concluded from previous researchers' work [1]. This characteristic would become even more demanding if it was to be considered in healthcare facilities such as hospitals where health-defected populations might be exposed to external sources of microbes and it is a subject of concern for international health organizations such as World Health Organization (WHO) [2].

***Corresponding author.**
*E-mail address: mostafaessameissa@yahoo.com

Published by Yıldız Technical University Press, İstanbul, Turkey

Biological stability is one of the important properties of the municipal water distribution system as could be demonstrated by Gauthier et al. [3]. This stability ensures the predictable quality and consistency of the delivered water to the final consumers as could be demonstrated according to WHO (2017) [4]. Saha and Donofrio work showed a microbiological study of the metalworking which might lead to a concern that any uncontrolled excursions in the bioburden could lead to undesirable consequences, especially with significantly variable microbiological properties [5]. Several control techniques such as exposure to Ultraviolet (UV) irradiation and application of biocidal agents have been used [5, 6].

When considering a large amount of data gathered for microbiological trends of water distribution networks, the implementation of the multivariate analysis should be considered to analyze the biological quality of this type of complex system. Principal component analysis (PCA) is a multivariate data analysis tool commonly used to reduce the dimension (number of variables) of many interconnected variables while retaining as much information (variation) as possible. PCA calculates a set of uncorrelated variables (components or PCs). These components are ordered so that the first few hold the most variability that is present in all the original variables [7–10].

The present study aimed to investigate the microbiological stability of municipal water distribution network systems using a multivariate analysis approach. The study focused on the use of total bioburden count parameter in water examination as an indicator for the biological stability of city water. This work investigated the variability and the association in the microbiological quality between different sections of the water distribution system. Based on the outcome of the research herein, further investigation might be extended to cover other quality inspection criteria such as objectionable microorganisms' trend with dissemination profile, Total Organic Carbon (TOC) and conductivity.

MATERIALS AND METHODS

Long-term data collection – for 45 successive months - was performed for regular monitoring of the investigated subject for Multivariate Analysis (MVA) study following a discussion by Great Learning Team [11].

Subject Study

Healthcare facility with distribution water network system that gained its feed supply of city water from a municipal line was studied following the method described by Ainsworth [12]. The distribution of the piping network supplied different compartments in the plant and each section has its own sampling port from which routine samples of city water were collected for analysis.

Collection of Samples

Water samples were collected aseptically in sterile bottles from each functioning point-of-use after prior allowance for appropriate flushing as described by Essam Eissa and EMSL Analytical [13, 14]. Well-closed sample bottles were submitted for analysis immediately within two hours so that the aseptically collected samples would be analyzed freshly as soon as possible to avoid any variations from the original microbial count. The guidance provided with appropriate modifications were done following United States Environmental Protection Agency (USEPA) (2016) [15]. Storage of water bottles in refrigerator (2–8°C) was ensured for about eight hours, if samples would not be treated at once in the laboratory [15].

Microbiological Testing of Water Samples

Water samples were tested microbiologically using aseptic tools and techniques to avoid extraneous contamination. Sterile, single-use and 0.45 µm disposable membrane filters were used to collect microorganisms in a definite volume of water sample through a vacuum filtration system where filters were transferred aseptically over agar media for detection of coliforms and enumeration of the Total Viable Aerobic Count (TVAC) following method given by Eissa (2018a) [13]. The plates were incubated and counted for determination of the number of Colony Forming Unit (CFU) per unit volume of water and for detection of the characteristic appearance of the pathogen colonies. Francy et al. and Ngwa et al. [16, 17] had provided guidance for microbiological testing of water samples. Confirmatory biochemical tests were conducted for any suspicious colonies using appropriate identification kits such as API following Gram staining technique.

Data Collection and Processing

All observations of the results were recorded in the columns of the worksheets of commercial statistical software. Goldwater demonstrated the use of Excel program for data analysis [18]. Datasets were arranged in chronological order with all use points identification codes assigned for each column in GraphPad Prism program [19]. Each sampling port had a "C" letter followed by its serial number for location identification i.e. C1, C2, C3, ... etc. The processed database was subjected to statistical analysis and multivariate study using Principal Component Analysis (PCA) as could be described earlier by Eissa and colleagues [20].

RESULTS AND DISCUSSIONS

The present study covered the long-term trend monitoring of the total microbiological plate count of the city water in a limited location of the distribution network system in a selected healthcare facility over 45 months of regular sampling and analysis which was done as previously performed by similar work in other circumstances [13]. No objection-

Table 1. Eigen analysis of the Correlation Matrix for sampling ports from C1 to C13 of water distribution system

PCA of Correlation-Type Matrix¥			
Component number€	**Latent roots£**	**Proportion**	**Cumulative**
1	6.0278	0.464	0.464
2	1.5004	0.115	0.579
3	1.1142	0.086	0.665
4	1.0200	0.078	0.743
5	0.8159	0.063	0.806
6	0.6216	0.048	0.854
7	0.5429	0.042	0.896
8	0.3502	0.027	0.923
9	0.2740	0.021	0.944
10	0.2477	0.019	0.963
11	0.2156	0.017	0.979
12	0.1651	0.013	0.992
13	0.1044	0.008	1.000

¥: Computation was done using Minitab® 17.1.0.0; £: Characteristic values: The variances of the principal components; € The components were studied as a linear combinations of the originally recorded variables.

Table 2. Multivariate analysis of municipal water network of the first three principal components (PCs)

Dimensionality reduction¥			
Variables€	**PC1**	**PC2**	**PC3**
C1	0.305	0.307	-0.136
C2	0.279	0.347	-0.171
C3	0.303	0.347	-0.171
C4	0.172	0.332	0.693
C5	0.304	-0.297	0.214
C6	0.262	-0.254	0.470
C7	0.304	-0.171	-0.132
C8	0.333	-0.030	-0.331
C9	0.253	-0.116	0.178
C10	0.301	-0.263	-0.049
C11	0.307	-0.286	-0.131
C12	0.303	0.065	-0.141
C13	0.054	0.373	0.065

¥: Calculation was done using Minitab® 17.1.0.0; €: PCA was used to reduce the datasets into a smaller number of components.

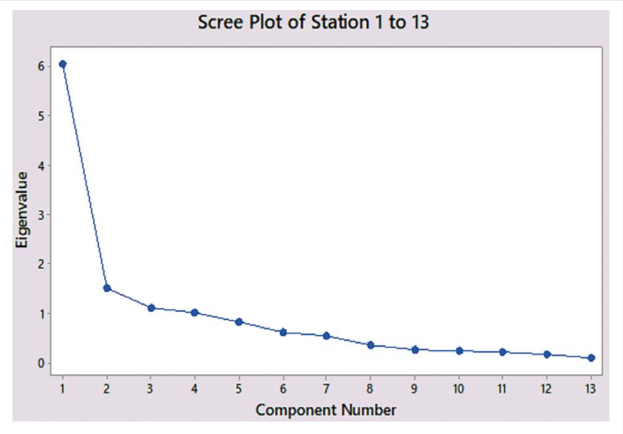

Figure 1. Scree diagram showing that the first five components could explain more than 80% of data variability.

able or pathogenic microorganisms were detected which followed the same working principle as done by Essam Eissa but with different outcome [21]. A low level of the microbial count was observed in most samples with all representative use points – for the different sections of water lines – were significantly (p=0.05) lower in bioburden content than the reference of the maximum threshold acceptance criterion of 100 CFU/mL which was similar in the outcome as in the previous work. A possible explanation for the observed low level of the microbial count could be attributed to the presence of the residual amounts of the sanitizing agent in the municipal water to deliver water with acceptable microbiological quality for human consumption.

Multivariate analysis using PCA in Table 1 showed Eigen values and the proportion of each component along with the cumulative contribution using the correlation-type of the matrix [22]. The first, second and third components contributed 46.4%, 11.5% and 8.6% of the variance that they could explain, respectively. Based on the selection for the Pareto principle of 60/40, the first three components would explain more than 66% of variations in the database [23]. Dimensionality reduction was ensured by including the highly influential and acceptable number of the principal components that could account for most of the variability [24]. With this respect, the fifth component was the first component number that yielded a characteristic value lower than unity with the cumulative contribution of the first five PCs above 80% from the total computed factors as could be seen in Figure 1. Based on the Kaiser criterion, PCs with latent values greater than one should be considered in the analysis [24]. Nevertheless, the scree graph started to form an almost straight line after the third component with a transition elbow curve between the second and the third components could be detected.

From the previous argument, the three first PCs would be considered in the present study for simplification and insight into data visualization. Table 2 showed the coefficient for each component calculated for each contributing variable i.e. use points representing different parts of the municipal water distribution system [24]. The greater the absolute fig-

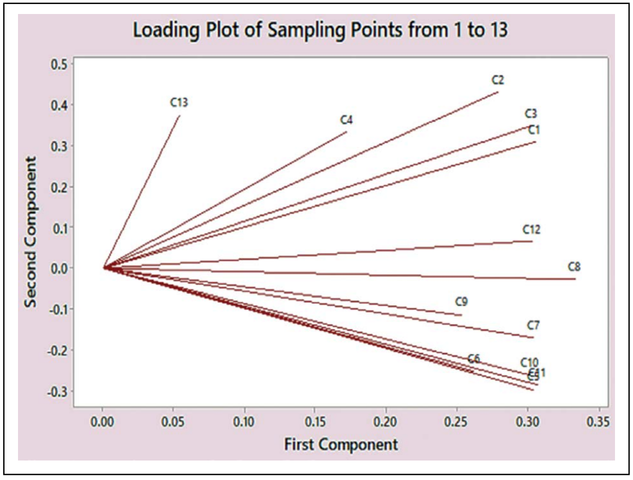

Figure 2. Loading diagram showing the influence of different points-of-use in the water distribution network on both the first and second components.

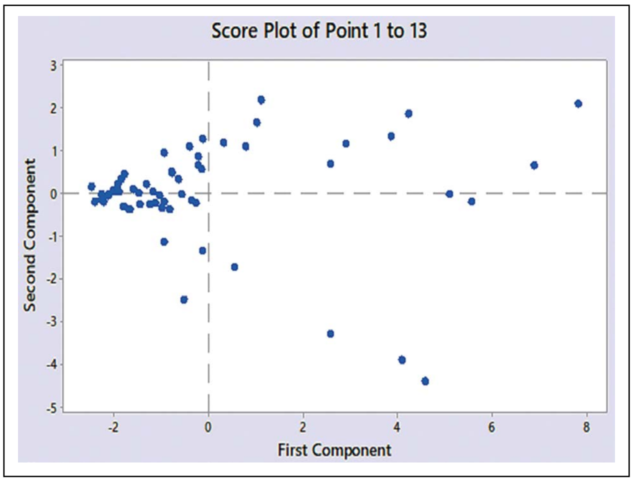

Figure 3. Score diagram showing data distribution, trend and pattern surrounding the reference zero point, in addition to the detection of outliers and clustering tendency.

Table 3. Aberrant values detection and removal statistically using ROUT method

	ROUT$^{¥}$ (Q=1.000%)**													
Method*	**Distribution points in facility**													**Overall network**
Number of points	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	Average
Analyzed	59	59	59	59	59	59	59	59	59	59	59	59	59	59
Outliers$^{€}$	12	6	12	8	9	11	8	10	6	11	13	8	10	7

*: Q is the maximum False Discovery Rate (FDR); **: Computation was done using GraphPad Prism for Windows version 6.01; ¥: Robust regression and Outlier removal combination technique; €: Aberrant values extracted from a non-linear regression.

ure of the coefficient, the more influential the corresponding water distribution section is in calculating that component. From Table 2, the descending order for the effect of points-of-use on the first component was as the following: C8, C11, C1, C5, C7, C3, C12, C10, C2, C6, C9, C4 and C13. Concerning the second component, the decrease in the coefficient was as the following: C13, C3, C2, C4, C1, C12, C8, C9, C7, C6, C10, C11 and C5. Finally, the last component showed the following descending order: C4, C6, C5, C9, C13, C10, C11, C7, C1, C12, C3, C2 and C8. A more in-depth focus on the first two components was desired since these two PCs only accounted for about 58% of the variability in the database of the total microbiological water plate count.

The loading plot in Figure 2 showed the contribution of each section of the municipal water distribution network on the first two principal components that demonstrated the main impact on the data variability of microbiological plate count [25]. The first component appeared to be affected by the microbiological density expressed as CFU/mL where all points were located at the positive side of the x-axis (PC1). Accordingly, monitoring ports C13 and C4 showed a lower trend in HPC. PC2 likely added another dimensionality of the relative locations for different lines with respect to the feeding source [25]. Thus, the microbial burden in water over the length of the water line might determine a positive quality (+y-axis) versus negative municipal water quality (-y-axis), a marker that could be used as one of the indicators for microbiological stability through the facility network system. This marker would explain bioburden quality as total microbial count over distance from the reference entry point of the distribution water section. Moreover, the graph demonstrated correlated points at minor and gross levels [26]. Normally, the biological quality of the feeding source influences that of the distribution network. However, at a smaller scale the microbiological quality level would be affected by the flow pattern of water in different sections, in addition to the state of the distribution system at different affected sections ex. corrosions. Other factors such as hygiene level, microbial sources and human population density cannot be ruled out as possible influential factors in the observed microbial trend. At a gross level, three groups were identified C1→C4, C5→C12 and solitary C13 that stands alone. On the other hand, small-scale correlation showed a closer association between C5. C6, C10 and C11 following GraphPad Software LLC guiding manual [27]. In the same line, a binary closer correlation was found between C1 and C3, in addition to C7 and C9.

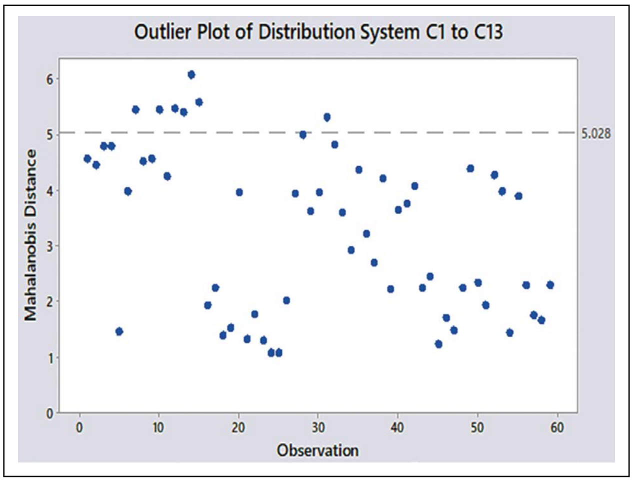

Figure 4. Outlier diagram showing outlier values due to unusual high values in microbiological count data.

The score plot was used as a scatterplot charting for the first component against the second one. There was a shred of strong evidence for clustering tendency that could be observed in the left side of the y-axis in the score plot in Figure 3. Moreover, outlier values could be expected also as some remote and isolated values were spotted toward the edges at the right position above and below the +x-axis [20, 26, 28]. Sporadic high values in the database records were detected, despite no Out-Of-Specification (OOS) results could be detected as all results were below the regulatory acceptance specification limit of 100 CFU/mL. In addition, Points were not randomly dispersed in the graph [25]. Accordingly, data did not show the expected behavior of the normal distribution spreading with possible skewness would be expected due to solitary excursions in the microbial count of water samples.

Since aberrant values could be spotted in the preceding section, the presence of excursion would be analyzed. Possible reasons for such abnormal results include the quality state of the piping system and the nature of water circulation and flushing, in addition to the design geometry. Also, the presence of dead ends and/or stagnant spots cannot be ruled out. A preliminary analysis for detection of the excursions was done using ROUT method – at recommended Q=1.000% - in GraphPad Prism version 6.01 [29]. The results obtained in Table 3 showed the following ascending order of the outlier rates (expressed as percent): (C2, 10.17%), (C9, 10.17%), (C4, 13.56%), (C7, 13.56%), (C12, 13.56%), (C5, 15.25%), (C8, 16.95%), (C13, 16.95%), (C6, 18.64%), (C10, 18.64%), (C1, 20.34%), (C3, 20.34%) and (C11, 22.03%). The rate of the emergence of the aberrant values for the mean of all distribution points for the overall facility assessment was 0.1186. The outlier plot displayed the Mahalanobis distance in Figure 4 for all observations, in addition to a reference line to identify outliers. The Mahalanobis distance is defined as the distance separating each data point from the centroid of multivariate space (in layman's term: the overall mean) [25]. Examining Mahalanobis distances was viewed as a more powerful mean for discovering the outliers better than looking at a single variable at a time because it took into consideration the diverse scales between the variables and the correlations between them.

CONCLUSIONS

In conclusion, multivariate analysis was surfaced in this study when we aimed to measure the microbiological quality (in terms of TVC) of the city water within a specific facility. The first dimension was the degree of the stability of microbial count over a certain time for each water line in the plant. The outlier rate of the HPC results – even there was no OOS - over a specific period might indicate a state of abnormal variation in the quality of water which should be investigated and eliminated. The second one could be found through the PCA where the position of each segment of the water network showed considerable variation in the average total microbial count from the neighbor lines with clustering tendency was obvious indicating a significant association between the correlated points that formed grouping environment. The first PC demonstrated the imparted difference in the bioburden associated with each use point. Moreover, the tested sampling ports showed unique PC2 because of its expected link to the relative location of each point-of-use from the source feed to a great extent. No doubt that the original quality of the municipal water was basically affected by the feeding source from the supplementing city water station. However, it was out-of-scope in this study and would be evaluated in another future analysis. Nevertheless, the distribution network system within a specific facility had its influential contributing factor over water quality before reaching the consumers. Importantly, it was the methodology and the findings herein which could find their room in broader applications likewise such as different types of water including purified water and Water for Injection (WFI), in addition to the inspection of other quality characteristics of water such as Total Organic Carbon (TOC) and conductivity.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] M. D. Sobsey, C. E. Stauber, L. M. Casanova, J. M. Brown, and M. A. Elliott, "Point of use household drinking water filtration: A practical, effective solution for providing sustained access to safe drinking water in the developing world," Environmental Science & Technology, Vol. 42(12), pp. 4261–4267, 2008. [CrossRef]

[2] G. Ducel, J. Fabry and L. Nicolle, "Prevention of hospital-acquired infections," 2nd ed., World Health Organization, Geneva, 2002.

[3] V. Gauthier, B. Gérard, J. Portal, J. Block and D. Gatel, "Organic matter as loose deposits in a drinking water distribution system," Water Research, Vol. 33(4), pp. 1014–1026, 1999. [CrossRef]

[4] World Health Organization (WHO), "Potable Reuse: Guidance for producing safe drinking-water," 2017. https://apps.who.int/iris/bitstream/handle/10665/258715/9789241512770-eng.pdf Accessed on Jan 5, 2022.

[5] R. Saha and R. Donofrio, "The microbiology of metalworking fluids," Applied Microbiology and Biotechnology, Vol. 94(5), pp. 1119–1130, 2012. [CrossRef]

[6] M. E. Eissa, "Assessment for Sporicidal activity of two types of peroxygen/silver-based disinfectants: A comparative study," Journal of Pharmaceutical Technology, Research and Management, Vol. 4(2), pp. 161–169, 2016. [CrossRef]

[7] C. B. Cordella, PCA: The basic building block of chemometrics, In I. S. Krull, (Eds.). Analytical Chemistry," IntechOpen, London, United Kingdom, 2012.

[8] I. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 374(2065), Article 20150202, 2016. [CrossRef]

[9] S. Salata and C. Grillenzoni, "A spatial evaluation of multifunctional Ecosystem Service networks using Principal Component Analysis: A case of study in Turin, Italy," Ecological Indicators, Vol. 127, Article 107758, 2021. [CrossRef]

[10] M. E. Eissa, E. Rashed and D. E. Eıssa, "Quality improvement in routine inspection and control of healthcare products using statistical intervention of long-term data trend," Dicle University Journal of the Institute of Natural and Applied Science, Vol. 10(2), pp. 163–184, 2021.

[11] Great Learning Team, "Overview of multivariate analysis | what is multivariate analysis and model building process?," GreatLearning Blog: Free Resources what Matters to shape your Career!, 2020. https://www.mygreatlearning.com/blog/introduction-to-multivariate-analysis/ Accessed on Jan 5, 2022.

[12] R. Ainsworth, "Safe piped water: Managing microbial water quality in piped distribution systems," Water Intelligence Online, Vol. 12, 2013. [CrossRef]

[13] M. Essam Eissa, "Investigation of microbiological quality of water from the feed source to the terminal application in the healthcare facility: A case study," Health Research, Vol. 2(1), pp. 16, 2018. [CrossRef]

[14] EMSL Analytical, Inc, "Legionella, fecal coliform & total coliform sampling guide," Emsl.com, 2022. https://www.emsl.com/PDFDocuments/Sampling-Guide/legionella_coliform_guide.pdf. Accessed on Jan16, 2022.

[15] United States Environmental Protection Agency, "Quick guide to drinking water sample collection," Epa.gov, 2022. https://www.epa.gov/sites/default/files/2015-11/documents/drinking_water_sample_collection. pdf. Accessed on Jan15, 2022.

[16] D. Francy, E.A. Stelzer, A.M. Brady, C. Huitger, R.N. Bushon, H.S. Ip, M.W. Ware, E.N. Villegas, V. Gallardo and H.D. Lindquist, "Comparison of filters for concentrating microbial indicators and pathogens in lake water samples," Applied and Environmental Microbiology, Vol. 79(4), pp. 1342–1352, 2012. [CrossRef]

[17] G. Ngwa, R. Schop, S. Weir, C. León-Velarde and J. Odumeru, "Detection and enumeration of E. coli O157:H7 in water samples by culture and molecular methods," Journal of Microbiological Methods, Vol. 92(2), pp. 164–172, 2013. [CrossRef]

[18] E. Goldwater, "Using excel for data analysis," People.umass.edu, 2022. https://people.umass.edu/evagold/excel.html. Accessed on Jan 19, 2022.

[19] H. Motulsky, "Home - GraphPad", Graphpad.com, 2022. https://www.graphpad.com/ Accessed on Jan 18, 2022.

[20] M. Essam Eissa, "Determination of the microbiological quality of feed city water to pharmaceutical facility: distribution study and statistical analysis," Athens Journal of Sciences, Vol. 4(2), pp. 143–160, 2017. [CrossRef]

[21] M. Eissa, "Evaluation of microbiological purified water trend using two types of control chart," European Pharmaceutical Review, Vol. 23(5), pp. 36–38, 2018.

[22] P. Praus, "Urban water quality evaluation using multivariate analysis," Acta Montanistica Slovaca, Vol. 12(2), pp. 150–158, 2007.

[23] L. Almagro, X. Llabres and P. Cintas, "Industrial Statistics with Minitab," Hoboken, N.J., Wiley, 2013.

[24] W. Ojok, J. Wasswa and E. Ntambi, "Assessment of seasonal variation in water quality in River Rwizi using multivariate statistical techniques, Mbarara Municipality, Uganda," Journal of Water Resource and Protection, Vol. 9(1), pp. 83–97, 2017. [CrossRef]

[25] Minitab® 18 Support. 2019, “Interpret all statistics and graphs for Factor Analysis - Minitab", Support.minitab.com, 2022. https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/factor-analysis/interpret-the-results/all-statistics-and-graphs/ Accessed on Jan 12, 2022.

[26] D. Weragama, V. Weerasingha, L. Jayasumana, J. Adikari, J. Vidanarachchi and H. Priyashantha, “The physicochemical, microbiological, and organoleptic properties and antioxidant activities of cream cheeses fortified with dried curry leaves (Murraya koenigii L.) powder,” Food Science & Nutrition, Vol. 9(10), pp. 5774–5784, 2021. [CrossRef]

[27] GraphPad Software LLC, “GraphPad Prism 9 Statistics Guide - How to: Correlation,” Graphpad.com, 2022. https://www.graphpad.com/guides/prism/latest/statistics/stat_how_to_correlation.htm. Accessed on Jan 13, 2022.

[28] C. Charoux, L. Free, L.M. Hinds, R.K. Vijayaraghavan, S. Daniels, C.P. O'Donnell and B.K. Tiwari, “Effect of non-thermal plasma technology on microbial inactivation and total phenolic content of a model liquid food system and black pepper grains,” LWT, Vol. 118, Article 108716, 2020.

[29] M.E. Eissa, “Role of statistical process control of pharmaceutical product to monitor consistency of the manufacturing operation,” EC Pharmacology and Toxicology, Vol. 6, pp. 439–444, 2018.