Multivariate Modelling of Water Quality Parameters in Nigeria

Adesina O.A.[1], Onatunji A.P.[2*], Akinlade Y.O[3], Ogunjobi E.O.[4]

[1,3]Department of Mathematics and Statistics,The Polytechnic, Ibadan, Oyo State

[2]LAUTECH International College, Ogbomoso, Oyo State

[4]Department of Mathematics and Statistics, Adeseun Ogundoyin Polytechnic, Eruwa, , Oyo

*Corresponding Authors: ***Onatunji A.P.,*** *LAUTECH International College, Ogbomoso, Oyo State*

Abstract: *Inadequate safe water remains a threat to human health in Ibadan Metropolis of Oyo State, Nigeria. The Asejire reservoir is considered a major source of water in Ibadan and water pollution is pronounced as a result of growing population and urbanization in the area it serves. In this study, We studied and analysed water quality parameters of treated and untreated water samples from Asejire Reservoir of Oyo state, Nigeria. We examined the conformance of the sample parameters to the W.H.O 2017 standards for safe and acceptable drinking water parameters. Data were obtained from the Oyo State water corporation which contained values on water quality parameters such as turbidity, colour, pH, alkalinity, etc. Correlations between values of the pollutants were examined for collinearity before estimating linear discriminant functions (LD) that helped to classify water samples into any of "safe and acceptable", "safe but unacceptable", "unsafe yet acceptable" and "unsafe and unacceptable". The estimated discriminant function had a efficiency of up to 84.3% in correctly predicting the class of water sample based on "appearance", "colour", "turbidity", "taste and odour", "alkalinity", "hardness", "chloride", "iron", "manganese" and "total dissolved solids". Results also showed that of all the 41 water samples which were regarded as final and ready for distribution; 7.3% were potentially unsafe and unacceptable for consumption as they were either polluted with iron or dissolved solids from domestic, industrial, agricultural wastes, insecticides and pesticides.*

Keywords: *W.H.O., Safety, Acceptability, Standards, Linear discriminant function.*

1. INTRODUCTION

Water is a unique substance that has many physical properties distinct from those possessed by other liquid, gaseous, or solid materials existing on the earth's surface. Water is colourless and transparent substance which is not distributed uniformly over the surface of the earth, as some areas are blessed with a fairly uniform and more than adequate supply for human needs, while many other regions have a greater need for water than they can supply. Water is essential for the existence of all life forms; it is a vital need that man cannot survive without it. water is life and its importance in the life of man, animals and plants cannot be overemphasized. However, the task of meeting domestic water needs in rural and urban areas in most developing countries, particularly in Nigeria, is enormous and falls mainly to women and children. Households also spend considerable time and effort fetching water from sources such as rivers, streams, ponds, wells and boreholes. The most important fresh water resources are rivers (Kolawole et al. 2011). Rivers play a basic role in assimilating the urban waste water, industrial wastes and surface run off from agricultural fields (Basu and Lokesh 2014). Human being and other living organisms depend on water for their survival.Quality of different water resources is subjected to ongoing consequences of water pollution and these results in the increase in demand for monitoring its quality. Water is considered polluted if some substances or condition is present to such a degree that the water cannot be used for a specific purpose (W.H.O., 2017). If the concentration of substances naturally present in water increases then also the water is said to be polluted. Olaniran (1995) defined water pollution to be the presence of excessive amounts of a hazard (pollutants) in water of water quantity in such a way that it is no long suitable for drinking, bathing, cooking or other uses. Testing of the water quality is an essential part of environmental monitoring. The aquatic life as well as surrounding ecosystem is affected when water quality is poor. The quality must be tested with different physico-chemical parameters and the selection of parameter for testing

should depend on the water uses, quality and purity (Ijesird,2015).The variety technologies for treating water at the point use which are common methods are water filters, aeration and disinfection which are employed to remove the bacteria and pathogens that contaminate water to decrease the waterborne diseases. Access to clean water for drinking cooking, bathing and other household needs is fundamental. Proper household water and sanitation pracices can increase

resilience to waterborne disease risk. An adequate supply of safe drinking water is one of the major pre-requisites for a healthy life, while the problems of inadequate safe drinking water are of national concern in Nigeria. Physical test is applicable for physical appearance such as color, temperature, pH, turbidity, Total Suspended Solids (TSS) and Total Dissolved Solids (TDS). Chemical test is performed for Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), alkalinity, dissolved oxygen, and hardness. According to Tiwari (2015), only developed countries managed to monitor these criteria due to the availability of sophisticated analytical instruments, technology and trained manpower.

2. METHODOLOGY

2.1. Linear Discriminant Analysis for Observations with Two Classes.

Consider a set of observations (or vector) X (also called features, attributes, variables or measurements) for each sample of an object or event with known class Y. This set of samples is called the training set. The classification problem is then to find a good predictor for the class Y of any sample of the same distribution (not necessarily from the training set) given only an observation $x \in X$ (Venables and Ripley, 2002).

LDA approaches the problem by assuming that the conditional probability density functions $p(x|y = 0)$ and $p(x|y = 1)$ are both normally distributed with mean and covariance parameters (μ_0, Σ_0) and (μ_1, Σ_1), respectively. Under this assumption, the Bayes optimal solution is to predict points as being from the second class if the log of the likelihood ratios is bigger than some thresh old T, so that:

$$(X - \mu_0)^T \Sigma_0^{-1}(X - \mu_0) + \ln|\Sigma_0| - (X - \mu_1)^T \Sigma_1^{-1}(X - \mu_1) - \ln|\Sigma_1| > T$$

Without any further assumptions, the resulting classifier is referred to as QDA (quadratic discriminant analysis).

LDA instead makes the additional simplifying homoscedasticity assumption (*i.e.* that the class covariances are identical, so $(\Sigma_0 = \Sigma_1 = \Sigma)$ and that the covariances have full rank. In this case, several terms cancel:

$$X^T \Sigma_0^{-1} X = X^T \Sigma_1^{-1} X$$

$$X^T \Sigma_i^{-1} \mu_i = \mu_i^T \Sigma_i^{-1} X$$

becauseΣ_i is "Hermitian" $(a_{ij} = \overline{a_{ji}})$and the above decision criterion becomes a threshold on the dot product$w \cdot x > c$ for some threshold constant c, where

$$w = \Sigma^{-1}(\mu_1 - \mu_0)$$

$$c = \frac{1}{2}(T - \mu_0^T \Sigma_0^{-1} \mu_0 + \mu_1^T \Sigma_1^{-1} \mu_1)$$

This means that the criterion of an input x being in a class y is purely a function of this linear combination of the known observations.

In this work, we have used the Fisher's linear discriminant rule which maximizes the ratio between $SS_{between}$ and SS_{within}, and finds a linear combination of the predictors to predict group (Hardle and Simar, 2007).

2.2. Fisher's Linear Discriminant

Suppose two classes of observations have means μ_0 and μ_1 and covariances Σ_0 and Σ_1. Then the linear combination of features $w \cdot x$will have means$w \cdot \mu_i$ and variances$w^T \Sigma_i w$ for $i = 0, 1$. Fisher defined the separation between these two distributions to be the ratio of the variance between the classes to the variance within the classes:

$$S = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{(w \cdot \mu_1 - w \cdot \mu_0)^2}{w^T \Sigma_1 w + w^T \Sigma_0 w} = \frac{(w \cdot (\mu_1 - \mu_0))^2}{w^T (\Sigma_0 + \Sigma_1) w}$$

This measure is, in some sense, a measure of the signal-to-noise ratio for the class labelling. It can be shown that the maximum separation occurs when

$$w \propto (\Sigma_0 + \Sigma_1)^{-1} (\mu_1 - \mu_0)$$

When the assumptions of LDA are satisfied, the above equation is equivalent to LDA.

2.3. Multiclass LDA

In the case where there are more than two classes, the analysis used in the derivation of the Fisher discriminant can be extended to find a subspacewhich appears to contain all of the class variability.This generalization is due to Rao (1948). Suppose that each of C classes has a mean μ_i and the same covarianceΣ. Then the scatter between class variability may be defined by the sample covariance of the class means

$$\Sigma_b = \frac{1}{C} \sum_{i=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^T$$

Where μ is the mean of the class means. The class separation in a direction w in this case will be given by

$$S = \frac{w^T \Sigma_b w}{w^T \Sigma w}$$

This means that when w is an eigenvector of $\Sigma^{-1} \Sigma_b$, the separation will be equal to the corresponding eigenvalue.

3. RESULTS OF THE ANALYSIS

Water quality parameters were obtained from the Oyo State Water Corporation in Nigeria. Data contained values on water quality parameters such as colour, appearance, turbidity, hardness, pH, etc. for 3 groups ("raw", "aerator" and "final"). The "final" water is that which has been treated and would be distributed to individual consumers.

Table1. *Conformance of Water sample parameters to W.H.O standard for safe andacceptable drinking water*

| Parameters	Observed Range	W.H.O standard (Health based)	W.H.O. standard (Acceptability)	Conformance of Final Water Sample	Period of Non-conformance of Final water sample
Turbidity	0 – 3.1	< 4 NTU	< 4 NTU	Yes	Nil
Color	5	< 15 TCU	< 15 TCU	Yes	Nil
Ph	7.0 – 8.2	6.5 – 8.5	6.5 – 8.5	Yes	Nil
Alkalinity	44.0 – 80.0	No health based guideline	< 100 mg//l	Yes	Nil
Total Hardness	50.0 – 104.0	< 500 mg/l	(100 – 300) mg/l for Calcium hardness	Yes	Nil
Chloride	12.0 – 34.0	No health based guideline	< 250 mg/l	Yes	Nil
Iron	0 – 25	< 0.3 mg/l	< 0.3 mg/l	No	March, 2014
Total Dissolved Solids	54 – 1402	< 600 mg/l	< 1000 mg/l	No	July, 2017; August, 2017
Manganese	0	< 0.4 mg/l	< 0.1 mg/l	Yes	Nil

Table 1 shows the acceptable standard units of measurement for all considered sample parameters in the study. Observed ranges and WHO standard health based and acceptability for water quality were

contained in the table.Here, a keen interest was on the conformance of the final water sample to the W.H.O standards for safe and acceptable drinking water.

.**Table2.** *Proportion of safe/unsafe final water*

	Acceptable	Unacceptable	Total
Safe	38 (92.7%)	0 (0.0%)	38 (92.7%)
Unsafe	0 (0.0%)	3 (7.3%)	3 (7.3%)
Total	38 (92.7%)	3 (7.3%)	41 (100.0%)

The table above gives a description of the final water sample classification into groups (safe/unsafe) according to the W.H.O drinking water standards. It shows the 92.7% with corresponding value of 38 of final water correctly classifed for safe and and acceptable for human comsumption while 7.3 % of final water correctly classified for unsafe and acceptable for human use. No proportion of final water sample was wrongly classified.
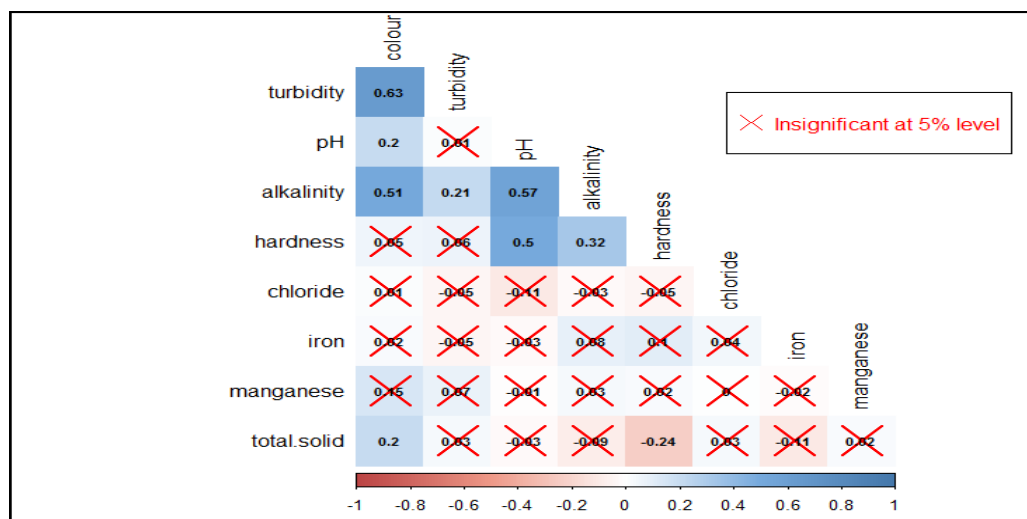
Figure1. *Correlation plots for water quality parameters*

Figure 1 contains Correlation plots for water quality parameters. From Figure 1 above, it is seen that variables such as colour, alkalinity, and total hardness correlated significantly with at least one other parameter. Hence, in order to make the model more parsimonious; we have removed colour, alkalinity and total hardness from the final set of variables (appearance and odour are categorical with groupings 1 and 2) used for the linear discriminant function.

Table3. *Prior Probabilities for water sample classified based on the WHO standard*

	Safe & acceptable	Safe but unacceptable	Unsafe yet acceptable	Unsafe & unacceptable	Total
Count (%)	62 (51.24%)	0 (0.0%)	2 (1.65%)	57 (47.11%)	121 (100.0%)

Table 3Each water sample ("raw", "aerator" and "final") was classified based on the W.H.O standard as either "1 - Safe and Acceptable", "2 – Safe but Unacceptable", "3 – Unsafe yet acceptable" and "4 – Unsafe and unacceptable". It contains the prior proballilies of safe/unsafe and acceptable/unacceptable water for domestic uses. There are 62, 0, 2 and 57 with corresponding percentages of 51.24, 0, 1.65 and 47.11 respectively for safe and acceptable, safe but unacceptable, unsafe yet acceptale and unsafe and acceptable. This implies that there is moderately high value of prior probability of 51.24%for safe and acceptable water. However, there is prior probability of 47.11 % of unsafe and unacceptable water.

Table4.ial*Coefficient of Linear discriminants*

Parameters	LD1	LD2
Appearance	-1.628	-0.592
Taste and Odour	-1.628	-0.592
Turbidity	0.023	-0.018
Alkalinity	-0.018	-0.003

Total Hardness	-0.013	0.005
Chloride	0.003	-0.008
Iron	0.100	-0.017
Manganese	-0.518	-10.359
Total Solid	0.001	-0.005
Proportion of Trace	**0.8916**	**0.1084**

Table 4 shows the coefficients of linear discrimanat functions. The table indicates variables that Alkalinity and Total Hardness have little discriminant ability for the estimation of safe and unsafe water sample for domestic uses in LD1, while appearance, taste and odour, chloride, iron, manganese and total solidcontrasted in the estimation . However, there is increase in number of sample parameters involved to have little discriminant ability for the estimation in LD2, namely, Turbidity, Alkalinity, Chloride, Iron and Total in this study. This implies that Alkalinity, Total Hardness in LD1 and Turbidity, Alkalinity, Chloride, Iron and Total solid have almost no discriminant power in comparism to other variables in both LD1 and LD2.

Table5. *Linear Discriminant model classification with respect to original groupings*

		Predicted		
		Safe &acceptable	Unsafe yet acceptable	Unsafe & unacceptable
Observed	Safe and acceptable	62	0	0
	Unsafe yet acceptable	0	2	0
	Unsafe and unacceptable	15	4	38

Table 5contains model classification with respect to original groupings. It reveals that 62, 2 and 38 are safe and acceptale, unsafe yet acceptale and unsafe and unacceptable respectively for original grouping of final water. This indicates that sixty two(62) observations were correctly classified for safe and acceptable water for human consumption with zero number of missclassified observation for other two groups . Two (2) and thirty eight (38) observations were classified corrrectly for "unsafe yet acceptable" and"unsafe and unacceptable" water respectively for human consumption. It also shows that 19 obervations misclassified in the model.

Table6. *Predicted posterior probabilities for safe and unsafe water*

		Predicted group		
		Safe and Acceptable	Unsafe yet acceptable	Unsafe and unacceptable
Observed	Safe and Acceptable	0.5041	0.000	0.0082
	Unsafe yet acceptable	0.000	0.0165	0.0000
	Unsafe and unacceptable	0.1157	0.033	0.3223

$Goodness\ of\ Fit\ =\ Sum\ \big(diag(predicted\ posterior\ probabilities)\big)$
$=\ 0.5041\ +\ 0.0165\ +\ 0.3223\ =\ 0.8429\ =\ 84.29\%$

Table 6 shows the predicted posterior probability of safe and unsafe water in Oyo state. The results obtained in the table are the probabililities of groups of water sample. "Safe and acceptable", "unsafe yet aceceptable" and "unsafe and acceptable" water with corresponding posterior probabilities in percenatage are 50.4, 0.02 and 32.0. This indicates that people in Oyo State have moderate high probability(0.5041) of drinking safe and acceptable water while the same people have low probability(0.0165) of drinking unsafe yet acceptable water in Ibadan area of Oyo State. In addition, th estimated water qaulity shows the goodness of fit with 84.29%, the estimated model is accurately fiited for prediction.

4. DISCUSSION

We studied and analysed water quality parameters of treated and untreated water samples from Asejire River of Oyo state, Nigeria. We examined the conformance of the sample parameters to the W.H.O 2017 standards for safe and acceptable drinking water parameters. Data were obtained from the Oyo State water corporation which contained values on water quality parameters such as turbidity, colour, pH, alkalinity, etc. Correlations between values of the pollutants were examined for

collinearity before estimating linear discriminant functions (LD) that helps to classify water samples into any of "safe and acceptable", "safe but unacceptable", "unsafe yet acceptable" and "unsafe and unacceptable". The estimated discriminant function had a efficiency of up to 84.3% in correctly predicting the class of water sample based on "appearance", "colour", "turbidity", "taste and odour", "alkalinity", "hardness", "chloride", "iron", "manganese" and "total dissolved solids".

5. CONCLUSION

From the study, the following conclusions were made:

1. That in March 2014; the "final" water sample was polluted with iron at about 25 mg/l as against the W.H.O standard of below 0.3 mg/l.

2. Dissolved Solid was in excess of the 600 mg/l recommended for safe drinking water although below the acceptability range of 1000 mg/l detectable by taste.

3. In August 2017 again; dissolved solid exceeded both safety and acceptability thresholds of 1000 mg/l which would be detectable by taste.

4. Of all the 41 water samples which were regarded as final and ready for distribution; 7.3% were potentially unsafe and unacceptable for consumption as they were either polluted with iron or dissolved solids.

5. The linear discriminant function derived from table was 84.3% effective in predicting the class of water samples from Asejire river based on turbidity, Appearance, taste, alkalinity, hardness, chloride, iron, manganese and total dissolved solid constituents.

RECOMMENDATION

Water samples particularly the final one should be assessed for conformance to safety and acceptability before distribution. Also turbidity should be maintained at less than 0.5NTU and average of 0.2NTU as recommended by W.H.O for large municipalities such as Ibadan.

REFERENCES

[1] Hardle W. and Simar L. (2007). Applied Multivariate Statistical Analysis. Springer Berlin Heidelberg. pp. 289–303.

[2] World Health Organisation (2017). Guidelines for drinking-water quality: fourth edition incorporating the first addendum Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

[3] Olaniran N.S. (1995). Environment and Health: An Introduction, in Olaniran, N.S. et.al (Ed) Environment and Health. Lagos. Micmillan Nig. Pub. Co for NCF, 34-151

[4] Venables W. N and Ripley B. D. (2002). Modern Applied Statistics with S. (4th ed.). Springer Verlag.

[5] Kolawole O. M., Kolawole T. A., Olayemi A. B and Okoh A. I (2011).Assessment of Water Quality in Asa River (Nigeria) and Its Indigenous Clariasgariepinus Fish. International Journal of Environmental Research and Public Health 8(11):4332-52

[6] BasuS. andLokesh, K. S. (2014). Critical analysis of existing economic tools available for assessing river water quality. International Journal of Water Resources and Environmental Engineering, 6(11), 287-294.

[7] Rao R. C. (1948). The utilization of multiple measurements in problems of biological classification. Journal of the Royal Statistical Society, Series B. **10** (2): 159–203. JSTOR 2983775.

[8] IjesirdT.S.(2015).Water Quality Parameters –A Review. International Journal of Engineering Science Invention Research & Development; Vol. I Issue IX . www.ijesird.com e-ISSN: 2349-6185