DOI: http://doi.org/10.4038/sljastats.v20i1.8004

Sri Lankan Journal of Applied Statistics, Vol. (20-1)

Hierarchical Clustering to Determine the Toxicity Levels of Factory Workers

G. A. Dawodu[1*]

[1]Department of Statistics, College of Physical Sciences (COLPHYS), Federal University of Agriculture, Abeokuta (FUNAAB.), Ogun State, Nigeria.
[*]Corresponding Author: dawoduga@funaab.edu.ng

Received: 29th May 2019 / Revised: 11th September 2019 / Published: 10th December 2019

ABSTRACT

Miners, explorers, energy station workers, artisans, factory workers, school pupils and even mere roadside pedestrians and hawkers are exposed to one harmful chemical element or the other. These may be solid, liquid or gaseous in their natural states. A prudent individual will obviously want to determine the pollutants he/she is often exposed to in the course of executing his/her routine daily activities and check, amongst other things, his/her degree of exposure, the resultant toxicity level and consequently determine ameliorative measures that will help reduce the pollution before it causes serious ailments. This work purports to provide a very cheap and simple procedure (i.e. before or after the placement of safety gargets) that can be used to monitor toxicity levels of individual factory workers with a view to determining those (i.e. the highest toxicity level cluster) workers that the company must provide additional intensive medical care for now or even retire before things get even worst (i.e. their toxicity levels become lethal). A data on Artisans or factory workers is used to demonstrate the functionality of the technique as well as its simplicity.

Keywords: Agnes, Euclidean, Factory management, Samples, Hierarchical clustering, Manhattan, Toxicity level.

1. Introduction

In most countries of the world, there are illiterates or those that are not literate enough to be employable in “white-collar” establishments or well ventilated offices. Total lack of properly ventilated offices is also noticeable in institutions of higher learning, most especially in developing countries of the world. For instance, some universities, polytechnics; colleges of education still use Blackboard, chalk and duster during the course of lectures and academic discussions where whiteboard, marker, duster and projection of slides ought to have been utilized. Nursery and primary school kids still chew the tips of their lead-pencils; suck their water-color brushes etc. Even in very pollution and radiation conscious communities of the world, inhabitants cannot be totally free from exposure to say, carbon monoxide (e.g. as exhaust from silencers of automobiles) and naturally emitting radioactive elements (e.g. radon). However, the bulk of illiterates or unskilled laborers often engage in artisan (Dawodu *et al*., 2011), factory laborers work that exposes them to more dangerous substances (e.g. toxic waste, polluted water and gasses). Even skilled miners, explorers (mountain or cave) and energy station workers are equally exposed. The effects of pollution scourge cannot be totally estimated at a particular time, the best situation is to avoid it completely, the transmission of the scourge along a pedigree or lineage usually takes place through mutation and thus it effects may be noticed in children that were not even born when the pollution actually occur although in a minimal ebb. Factory managements, although are usually interested in the turnover and profit margins, almost all the time, but they too will not desire carrying the guilt of being the cause of death of many of their exposed workers. They are usually supposed to provide safety gargets and encourage their workers to use them while at work. In addition, the management is supposed to continue to research on ways to reduce the effects of the pollutants on the exposed workers throughout the existence of their factory and they will naturally be interested in less expensive ones so that such researches will not “eat” deep into their profit margins. The workers too will naturally be interested in less complicated and simple to understand techniques to monitor their levels of toxicity resultant from their exposure to the factory pollutants and toxic wastes. This work purports to provide a very cheap and simple procedure that can be used to do the said monitoring.

2. Materials and Methods

Data[1] was obtained from each volunteer, artisan or factory worker by measuring the quantities of Lead (Pb) (Babalola, *et al.*, 2010), Calcium (Ca), Sodium (Na), Potassium (K) and Magnesium (Mg) in the blood or urine samples of volunteers. As for atpase activities[2], two quantities were taken (i.e. Calcium-Magnesium Atpase Activity and Sodium-Potassium Atpase Activity). For plasma, only one was taken (i.e. Plasma-Calcium). Altogether, data on nine chemicals was taken on the volunteers. An extract of the data (Table 1) is shown below:

Table 1: An extract of the factory workers' data (the unit of measurement is ppm)

Workers Id	C1	C2	C3	C4	C5	C6	C7	C8	C9
W1	8.51	19.64	5.95	0.97	0.74	0.22	0.99	45.37	2.36
W2	6.1	18.45	3.77	0.62	0.5	0.12	0.8	23.05	2.19
W3	5.78	20.3	6.49	0.97	0.88	0.23	0.94	39.78	1.81
W4	12.1	16.58	6.93	2.18	1.02	0.29	1.38	47.53	1.84
W5	4.15	17.01	4.91	0.95	0.5	0.16	0.75	26.46	2.46

With the legend;

C1 denotes Lead in Urine
C2 denotes Lead in Blood
C3 denotes Erythrocyte Membrane Sodium
C4 denotes Erythrocyte Membrane Potassium
C5 denotes Erythrocyte Membrane Calcium
C6 denotes Erythrocyte Membrane Magnesium
C7 denotes Calcium-Magnesium Atpase Activity
C8 denotes Sodium-Potassium Atpase Activity
C9 denotes Plasma-Calcium
Wi denotes the identification of the ith worker, $i = 1, 2, \ldots, N$ [3]

[1] Data description varies from factory to factory, it depends on the chemicals (or elements) the artisans or factory workers are exposed to.

[2] Atpase activities are probable in this case as well but they may not be relevant if the toxic chemicals to which each factory worker is exposed do not present any atpase activities.

[3] N is either the population (or sample) size of the workers of the factory.

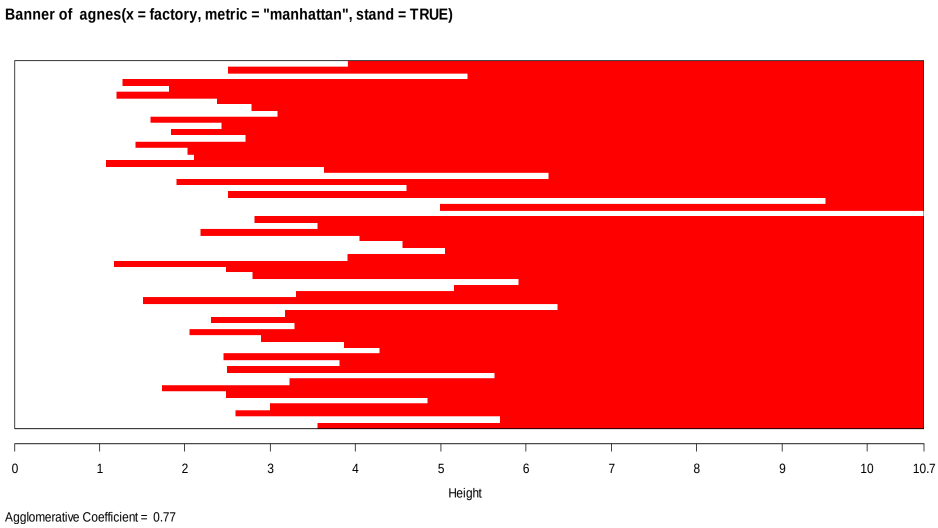
3. Results

The implemented package in R (Maechleret *al.*, 2016), known as **cluster** (Fraley and Raftery, 2002; Fraley *et al.*, 2009) that was written from its Fortran equivalent (Struyf *et al.*, 1996), contains all the subroutines for clustering (Sarkar, 2008, 2009; Lance and Williams, 1966). To illustrate how the package works, the data on sixty (60) factory workers or artisans that are exposed to the nine chemicals (section 2), was first categorized into four (i.e. lead in urine, lead in Blood, sum of Erythrocytes, sum of atpase activities and plasma activities) before supplying the data to agnes of the library "cluster", a package of R (Crawley, 2007) to obtain the following output (i.e. Figure 1).

```
Call:    agnes(x = factory, metric = "manhattan", stand = TRUE)
Agglomerative coefficient:  0.7655986
Order of objects:
 [1]  1  3 16  2 15 21 23 11 14 10 13 17 19 18 20 24 22 25 12  5  6  8  9  4  7 26 42
27 31 32 40 [32] 30 34 38 35 36 33 37 54 57 28 48 49 43 50 44 52 29 47 45 53 39
51 55 56 41 46 58 59 60

Height (summary):
   Min.  1st Qu.  Median   Mean  3rd Qu.   Max.
  1.078   2.243   2.896    3.421   4.166    10.668

Available components:
[1] "order"  "height" "ac"    "merge"  "diss"   "call"   "method" "data"
```



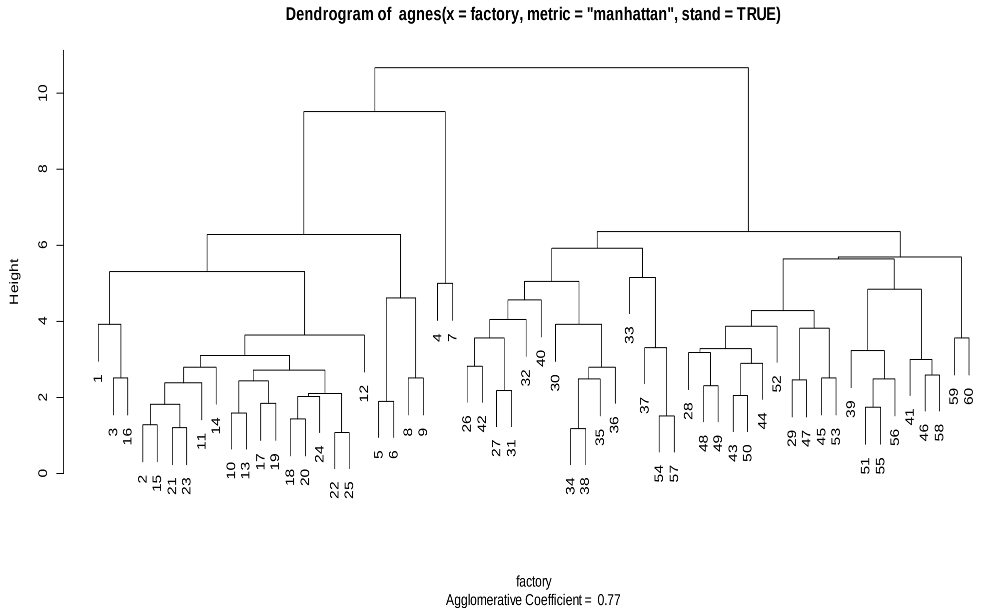


Figure 1: Result obtained when the program was ran on the data on factory workers.

4. Discussion & Conclusion

The increase in chemical toxicity in an individual factory worker is traceable in the Urine, Blood, Erythrocytes, Atpase activities, or Plasma, those are used here to create the clusters. As seen in the resultant dendrogram (Carmicheal and Sneath, 1969), in figure 1, there are, if we create clusters of size ten each, six clusters. The first cluster is {W1, W3, W16, …, W10} whilst the second is {W13, W17, W19, …, W5} and so forth (Kaufman and Rousseeuw, 2005). The readings for the first cluster of size 10 are:

Table 2: The readings for the first cluster of size 10
(i.e. from blood-lead to plasma-calcium)

19.64	8.51	5.95	0.97	0.74	0.22	0.99	45.37	2.36
20.3	5.78	6.49	0.97	0.88	0.23	0.94	39.78	1.81
17.84	8.51	5.95	1.24	0.62	0.23	0.79	35.38	1.86
18.45	6.1	3.77	0.62	0.5	0.12	0.8	23.05	2.19
18.11	8.29	3.82	0.97	0.7	0.16	0.69	23.8	2.27
12.44	5.11	4	0.84	0.68	0.17	0.67	24.36	2.16
12.06	9.51	3.49	1.31	0.54	0.15	0.65	25.95	2.01
7.4	1.65	4.16	0.7	0.59	0.18	0.71	21.33	2.24
23.46	12.11	3.52	1.06	0.65	0.18	0.64	30.53	2.16
16.85	2.18	5.11	0.93	0.75	0.17	0.76	27.8	2.34

The blood-lead (i.e. contents of the first column) gives the summary;

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
7.40	13.54	17.98	16.66	19.34	23.46

And standard deviation (sd), 4.712. Consequently, a 95% confidence interval is (7.236, 26.084) therefore a worker whose blood-lead reading is close to the upper limit, 26.084 (e.g. 23.46) should be removed from "high" exposure zones of the factory. This exercise can be repeated on the readings on other elements/chemicals, in other to know those that ought to be removed from the high exposure zones of those elements/chemicals. Also the exercise will be repeated for cluster 2 and so forth.

The usefulness in all of these will come handy when the management of the factory decides to, say, treat all their, say, lead (Pb) toxic workers, in batches of ten, then it will be expedient to keep the ordering being set in these clusters (Rousseeuw, 1986). The justification for this work is borne out of the fact that the technique is cheap, easy to; apply and comprehend. All that are needed are the; exposed factory workers, team of Biochemists (e.g. a large quantity of syringe and needles, required reagents and a fridge to preserve the samples), that will take samples and analyze to determine the quantities of the elements in them (i.e. in the workers' blood or urine in parts-per-million (ppm)) and an experienced environmental analyst or statistician that will do the coding to create the clusters. The import of all these is to help the managements of nuclear reactor; energy stations, factories involved with exposures to dangerous gasses (including naturally emitted and carcinogenic ones like radon etc.) to be aware of very cheap statistical tools that can be used to rank the exposures of their workers, most especially in this era of COVID-19. Before now, there was Severe Acute Respiratory Syndrome (SARS) and researchers are of the opinion that the combination of the duo to form Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) (Amanat and Krammer, 2020; Yang *et al.*, 2020) is the "probable" cause of COVID-19 pandemic. It is a global knowledge that respiratory track ailments are strongly linked with smoking and exposures to dangerous gasses (see better health and Mehrifar, Zamanian and Pirami, 2019). Managements wishing to pay "serious" attention to their workers' medical/welfare will find this work informative and immensely useful.

Acknowledgement

The Author is grateful to the entire members of staff of the Department of Biochemistry, in the defunct College of Natural Sciences (now Biosciences), FUNAAB, for their unflinching support during and after the numerous activities in the data collection and collation stages. The numerous volunteers spread around the consulted factory sites and artisans' workshops around Abeokuta (i.e. the study site) are highly appreciated too.

References

1. Amanat, F. and Krammer, F. (2020). SARS-COV-2 Vaccines: Status Report. *Immunity*, 52(4), 583-589.
2. Babalola, O. O., Okonji, R.E., Atoyebi, J. O., Sennuga, T. F., Raimi, M. M., Ejim-Eze, E. E., Adeniran, O. A., Akinsiku, O. T., Areola, J. O., John, O. O. and Odebunmi, S. O. (2010). Distribution of lead in selected organs and tissues of albino ratsexposed to acute lead toxicity. *Scientific Research and Essay*, 5(9), 845-848.
3. Carmicheal, J. W. and Sneath, P. H. A. (1969). Taxometric Maps. *Systematic Biology*, 18(4), 402-415.
4. Crawley, M. J. (2007). *The R Book*. John Wiley and Sons Ltd. The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
5. Dawodu, G. A., Asiribo, O. E., Adelakun, A. A., Ozoje, M. O., Ademuyiwa, O. and Akinwale, T. A. (2011). On the Vulnerability of the Blood of some Artisans' to Toxicity. *Journal of Environmental Statistics*, 2(4), 1-17.
6. Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97, 611-631.
7. Fraley, C., Raftery, A. E. and Wehrens, R. (2009). mclust: Model-based Cluster Analysis, URL http://www.stat.washington.edu/mclust, R package version 3.1-10.3.
8. Kantardzic, M. J. B. (2003). *Data Mining Concepts, Models, Methods and Algorithms*. IEEE Press Wiley-Interscience. A John Wiley and Sons, Inc.
9. Kaufman, L. and Rousseeuw, P. J. (2005). *Finding Groups in Data; An Introduction to Cluster Analysis*. John Wiley and Sons, Inc., Hoboken, New Jersey.
10. Lance, G. N. and Williams, W. T. (1966). A general theory of classificatory sorting strategies: Hierarchical systems, The *Computer Journal*, 9(4), 373-380.
11. Maechler, M., Rousseeuw, P., Struyf, A. and Hubert, M. (2016). Package cluster. Version 2.0.5. Maintainer: Martin Maechler maechler@stat.math.ethz.ch.
12. Mehrifar, Y., Zamanian, Z. and Pirami, H. (2019). *The International Journal of Occupational and Environmental Medicine*. 10(1), 40-49.
13. Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, New York, USA: Springer-Verlag.

14. Sarkar, D. (2009). lattice: Lattice Graphics. URL http://CRAN.R-project.org/package=lattice, R package version 0.17-22.
15. Struyf, A., Hubert, M. and Rousseeuw, P. J. (1996) Clustering in an Object-Oriented Environment. *Journal of Statistical Software*1.
16. Rousseeuw, P. J. (1986). *A visual display for hierarchical classification. Data Analysis and Informatics,* 4, North-Holland, Amsterdam. Pgs. 743-748.
17. Yang, Y., Xiao, Z., Ye, K., He, X., Sun, B., Qin, Z., Yu, J., Yao, J., Wu, Q., Bao, Z. and Zhao, W. (2020). SARS-COV-2: Characteristics and Current Advances in Research. *Virology Journal* (17), Article Number 117.