academicJournals

Vol. 6(3), pp. 26-34, March 2013
DOI 10.5897/AJMCSR 09.061
ISSN 2006-9731©2013 Academic Journals
http://www.academicjournals.org/AJMCSR

African Journal of Mathematics and Computer Science Research

Full Length Research Paper

Statistical grouping of cassava mosaic disease-resistant varieties cultivated by the National Root Crops Research Institute, Umudike, Nigeria

Nwabueze, Joy Chioma

Department of Statistics, Michael Okpara University of Agriculture Umudike, P. M. B. 7267 Umuahia, Abia State, Nigeria.

Accepted 4 March, 2013

Similarity level of secondary data on proximate composition of fufu flour measured from 44 cassava mosaic diseases (CMD) resistant varieties was obtained in order to group the 44 cassava varieties. The distant coefficient generated between the 44 CMD resistant varieties ranged from 0.000 to 51.510. The distance measures with low coefficient are grouped together. The agglomerative hierarchical schedule showed that the varieties could be placed in five distinct groups. The dendogram of the study which showed the relative size of the proximity coefficient at which cases were combined was obtained.

Key words: Similarity level, agglomerative, hierarchical schedule, proximity coefficients, proximate composition, cassava mosaic diseases (CMD)-resistant varieties.

INTRODUCTION

Nigeria is one of the largest producers of cassava in the world. Its production is currently put at about 34 million metric tonnes a years (FAO, 2002). Total area harvested of the crop in 2001 was 3.125 million harvests with an average yield of 10.83 tonnes per harvest. Presently cassava is primarily produced for food especially in form of *garri*, *lafun* and *fufu* with little or no use in agribusiness sector as an industrial raw material. The crop can be processed into several secondary products of industrial market value. These products include chips, pellets flour, adhesives, alcohol and starch, which are vital raw materials in the livestock feed, alcohol/ethanol, textile, confectionary, wool, food and soft drink industries. They are also traceable in the international market (FAO, 2004).

As a result of increase in cassava production in Nigeria, many improved cassava varieties have been discovered. The cassava varieties were bred for high yield, pest /disease resistance good product quality and early maturity. Improved cassava varieties for pest disease resistance are those improved cassava varieties capable of resisting the attack of common cassava disease known as cassava mosaic disease (CMD), a viral disease transmitted by a white fly vector (IITA, 2005). Researches have been going on for alternative uses of these improved cassava varieties. Today cassava has become an important biofuel crop in addition to its traditional role, its value as a fuel commodity.

In view of these, this study analyses a set of data reported by Etudaiye et al. (2009) on proximate composition of *fufu* flours processed from 44 CMD-resistant varieties cultivated by the National Root Crops Research Institute, Umudike, Nigeria. The data are subjected to cluster analysis as a multivariate technique for detecting natural grouping with the basic objective of reducing a large set of variables to a more meaningful smaller set (Crawford and Lomas, 1980). The approach is to put similar objects in the same group. Friedman and

E-mail: teeubueze@yahoo.co.uk. Tel: +2348068164190.

Rubin (1967) have adopted the view that cluster represents mixtures of multivariate normal population as a routine and basis for the design and clustering algorithm. Cunningham and Ogilvie (1972) adopted the concept of uttrametric space as a basis for the formation of cluster structure. It is expected that the algorithm will implore these space to recover cluster structure. Everitt (1993) applied the strong usual or spatial appeal to certain bivariate normal population mixtures to certain bivariate normal population mixtures to obtain several two dimensional plots. In this study, agglomerative hierarchical technique which usually produces dendrogram is employed. The principal difficulty with cluster analysis is that it introduces a concept of 'type' which is unscientific. As a result, we become embroiled with character-sketches and pen-pictures more appropriate to literature than to research. However, when the cluster analysis is understood and accepted, the technique can nonetheless serve an important heuristic function, provided certain conditions are met. This study among other objectives seeks to identify the most important nutritional factor in *fufu* flours from the 44 cassava varieties with the view to grouping them so that the varieties with similar nutritional composition will be in the same group.

EXPERIMENTAL DESIGN

Data used for the study

Secondary data on proximate composition of *fufu* flours processed from 44 different CMD-resistant varieties (Etudaiye et al., 2009) were used for this work. The amount in percentage of the proximate compositions of *fufu* flour made from each of these cassava varieties was measured and recorded by Etudaiye et al. (2009). These measurements include moisture (X_1), protein (X_2), ash (X_3), fat (X_4), fiber (X_5), carbohydrate (X_6) and dry matter (X_7) contents (Table 1).

Theoretical framework

Given the items and measurements, the basic requirement of cluster analysis include a quantitative scale for association between object which represent a measure of distance based on similarity or proximity, a clustering criterion and an applicable algorithm. Agglomerative hierarchical techniques which normally produce dendrogram were used in the study. This method starts with the calculation of the distance of each individual to all other individual groups are then formed by a process of agglomeration which is a process where all objects are placed alone in group of one, close groups are then gradually merged until finally all individuals are in a single group. The data usually consist of the value of *p* variable X_1, X_2…. X_p for *n* objects. These values are then used to produce an array of distance between the varieties given as:

$$d_{ij} = \sqrt{\sum_{k=1}^{n} \left(X_{ik} - X_{jk}\right)^2} \qquad (1)$$

Where X_{ik} is the value of variable X_k for variety I and X_{jk} is the value of the same variable for variety *j* and d_{ij} is the distance between variety *i* and variety *j*. These distances were then used for grouping.

RESULTS AND DISCUSSION

Cluster observations were executed in this work, subsequently the similarity levels were examined in order to determine the number of clusters to be used in the subsequent analysis. Cluster analysis is the organization of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based on similarity (Jain et al., 1999). The distance measures with low coefficient are grouped together. A variety of distance measures are in use in the various communities (Anderberg, 1973; Jain and Dubes, 1988; Diday and Simon, 1976). The agglomeration schedule showed that the varieties could be placed in five groups. The distance coefficient generated between the 44 cassava varieties ranged from 0.000 to 51.510. The hierarchical tree diagram or dendrogram showed the relative size of the proximity coefficient at which cases were combined. The operation of a hierarchical clustering algorithm produced a nested series of partitions based on a criterion for merging or splitting clusters based on similarity. Partitioned clustering algorithms identify the partition that optimizes a clustering criterion (Jain et al., 1999).

Using similarity level (S_L) of the amalgamation steps, Table 2 was obtained. This gave a simple distance measure like Euclidean distance (Table 2) which can be used to reflect dissimilarity between two patterns. Michalski and Stepp (1983) showed that other similarity measures can be used to characterize the conceptual similarity between patterns. From the amalgamation steps in Table 2, using squared Euclidean distance, Table 3 was obtained following the similarity level differences ($D = S_L - S_{L-1}$). An examination of the difference (D) in Table 3 revealed noticeable changes in amount at steps 33, 38, 40, 41 and 42 which suggest that 5 clusters would be appropriate. The number of clusters summarized in Table 4 is demonstrated in dendrograms obtained from the single-link algorithm (Jain and Dubes, 1988) (Figures 1 to 5). A dendrogram represents the nested grouping of patterns and similarity levels at which groupings change (Jain et al., 1999). The dendrogram can be broken at different levels to yield different clustering of the data. The clustering obtained demonstrated that the proximate composition of each CMD-resistant variety fall into several distinguishable clusters. The centroid of each of these clusters was determined by computing the mean of the moment vectors of the proximate composition falling into the cluster.

Cluster 1 has 3 observations (Figure 1), Cluster 2 has 15 observations (Figure 2), Cluster 3 has 23 observations (Figure 3), Cluster 4 has 2 observations (Figure 4) and

Table 1. Data on proximate composition of *fufu* flours from CMD-resistant varieties.

S/N	CMD varieties	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	97/4769	5.52	2.42	0.64	0.41	0.07	90.91	94.48
2	99/6012	6.17	1.93	0.64	0.24	0.06	91.38	93.83
3	94/0561	6.51	1.40	0.22	0.51	0.07	91.19	93.47
4	97/0162	6.97	2.07	0.32	0.58	0.05	90.06	93.03
5	94/0026	6.99	1.40	0.24	0.56	0.04	90.71	93.01
6	96/1642	7.02	2.55	0.39	0.04	0.04	90.60	92.98
7	98/0510	7.26	1.82	0.07	0.27	0.05	90.53	92.74
8	98/0505	7.29	2.45	0.05	0.20	0.06	89.95	92.71
9	99/3037	7.45	2.15	0.59	0.42	0.01	87.61	92.55
10	98/2101	7.46	1.93	0.52	0.42	0.06	87.37	92.54
11	97/4763	7.48	2.41	0.37	0.23	0.05	89.46	92.52
12	99/26324	7.53	1.40	0.31	0.44	0.06	89.62	92.47
13	97/2205	8.19	1.40	0.34	0.40	0.05	89.62	91.81
14	98/1565	8.19	0.70	0.51	0.34	0.07	89.50	91.81
15	92/0325	8.22	1.75	0.10	0.12	0.01	89.50	91.78
16	92B/00068	8.32	1.23	0.20	0.66	0.03	88.67	91.68
17	TME 419	8.32	0.70	0.41	0.39	0.06	90.12	91.68
18	96/0603	8.37	0.98	0.22	0.39	0.08	89.96	91.63
19	98/2226	8.62	2.80	0.45	0.34	0.02	87.76	91.38
20	92/0058	8.70	1.78	1.58	0.34	0.02	87.58	91.30
21	97/0211	8.70	2.80	1.53	0.56	0.04	86.19	91.30
22	95/0289	9.09	1.78	0.22	0.42	0.02	88.47	90.91
23	92/0326	7.13	2.80	0.34	0.56	0.08	87.09	90.87
24	4(2)1452	7.14	1.40	1.28	0.64	0.02	87.52	90.86
25	98/0002	9.32	1.78	0.43	0.41	0.04	88.01	90.68
26	97/4779	9.35	1.10	0.97	0.42	0.04	87.12	90.65
27	96/1642	9.35	2.80	0.48	0.56	0.01	86.80	90.65
28	96/0523	9.43	1.21	0.44	0.17	0.04	89.71	90.57
29	M98/0068	9.53	1.05	0.20	0.17	0.04	89.01	90.47
30	M98/0028	9.55	0.35	0.37	0.44	0.03	89.26	90.45
31	96/1089A	9.97	2.80	0.94	0.43	0.06	86.21	90.43
32	95/0166	9.63	2.55	2.13	0.03	0.03	86.18	90.37
33	92/0057	9.66	1.05	0.57	0.43	0.01	88.28	90.38
34	96/1314	9.78	2.10	3.34	0.29	0.05	86.93	90.22
35	97/3200	9.78	1.40	0.17	0.36	0.07	88.22	90.22
36	98/0040	9.96	1.05	0.53	0.42	0.05	87.99	90.04
37	TM/530572	10.13	1.75	0.77	0.33	0.03	87.32	89.87
38	99/2123	10.26	2.80	0.42	0.53	0.08	85.91	89.74
39	92/0067	10.38	1.08	0.77	0.60	0.04	87.16	89.62
40	97/0039	10.39	2.10	0.52	0.50	0.06	86.06	89.61
41	95/0379	10.45	1.23	0.47	0.43	0.04	87.83	89.55
42	92B/0061	10.65	1.23	0.08	0.48	0.08	76.89	89.35
43	98/0581	11.26	1.26	0.53	0.02	0.04	76.89	88.74
44	96/1569	12.25	2.10	0.52	0.29	0.03	84.81	87.75
Total		387.32	76.84	28.37	17.06	1.99	3867.98	4012.68
Mean		8.8027	1.7464	0.6448	0.3877	0.0452	87.9086	91.1973

X_1 = moisture, X_2 = protein, X_3 = ash, X_4 = fat, X_5 = fibre, X_6 = carbohydrate, X_7 = dry matter contents. Source: Etudaiye et al. (2009).

Cluster 5 has 1 observation (Figure 5). The observation variables with their group ID are displayed in Table 5 after the varieties have been sorted. It was evident from Table 5 and Figure 1 that CMD varieties in Cluster 1

Table 2. Amalgamation steps using squared Euclidean distance.

Step	No of cluster	Similarity	Distance level	Cluster joined		New cluster	No of observations in new cluster
1	43	99.94	0.146	17	18	17	2
2	42	99.92	0.201	39	41	39	2
3	41	99.92	0.205	23	27	23	2
4	40	99.90	0.271	33	36	33	2
5	39	99.90	0.276	13	15	13	2
6	38	99.86	0.362	22	25	22	2
7	37	99.84	0.417	8	11	8	2
8	36	99.83	0.441	14	17	14	3
9	35	99.82	0.468	5	7	5	2
10	34	99.82	0.483	24	26	24	2
11	33	99.81	0.505	33	35	33	3
12	32	99.79	0.564	4	6	4	2
13	31	99.78	0.593	28	29	28	2
14	30	99.76	0.636	38	40	38	2
15	29	99.74	0.684	37	39	37	2
16	28	99.70	0.798	2	3	2	2
17	27	99.56	1.162	42	43	42	2
18	26	99.54	1.209	28	30	28	3
19	25	99.53	1.238	12	13	12	3
20	24	99.48	1.384	4	5	4	4
21	23	99.37	1.661	23	31	23	3
22	22	99.29	1.870	32	34	32	2
23	21	99.28	1.898	20	24	20	3
24	20	99.24	2.001	12	14	12	6
25	19	99.21	2.089	19	22	19	3
26	18	99.05	2.501	12	16	12	7
27	17	99.02	2.598	21	23	21	4
28	16	99.00	2.644	33	37	33	6
29	15	98.79	3.189	4	8	4	6
30	14	98.74	3.328	1	2	1	3
31	13	98.55	3.834	9	10	9	2
32	12	98.24	4.643	19	20	19	6
33	11	97.14	7.552	21	38	21	6
34	10	96.65	8.858	28	33	28	9
35	9	96.61	8.951	4	12	4	13
36	8	96.15	10.175	19	28	19	15
37	7	96.00	10.577	21	32	21	8
38	6	93.53	17.097	4	9	4	15
39	5	92.98	18.560	19	21	19	23
40	4	90.97	23.864	1	4	1	18
41	3	84.54	40.835	19	44	19	24
42	2	57.57	128.103	1	19	1	42
43	1	0.00	264.204	1	42	1	44

(97/4769, 99/6012 and 94/0561) are low in moisture (X_1), average in protein (X_2) and ash (X_3), moderately high in ash (X_3) and fat (X_4) and high in carbohydrate (X_6) and dry matter (X_7) contents. Cluster 2 and Figure 2 show that CMD varieties 97/0162- 96/0603 are average with respect to moisture (X_1) and dry matter (X_7). The dendrogram of the cluster includes varieties 9, 10, 4, 6, 5, 7, 8, 11, 12, 13, 15, 14, 17, 18 and 16. Cluster 3 with

Table 3. Similarity level difference table.

S/N	S_L	S_{L-1}	D	S/N	S_L	S_{L-1}	D
1	99.94			23	99.28	99.24	0.04
2	99.92	99.92	0	24	99.24	99.21	0.03
3	99.92	99.90	0.02	25	99.21	99.05	0.16
4	99.90	99.90	0	26	99.05	99.02	0.03
5	99.90	99.86	0.04	27	99.02	99.00	0.02
6	99.86	99.84	0.02	28	99.00	98.79	0.21
7	99.84	99.83	0.01	29	98.79	98.74	0.05
8	99.83	99.82	0.01	30	98.74	98.55	0.19
9	99.82	99.82	0	31	98.55	98.24	0.31
10	99.82	99.81	0.01	32	98.24	97.14	1.1
11	99.81	99.79	0.02	33	97.14	96.65	0.49
12	99.79	99.78	0.01	34	96.65	96.61	0.04
13	99.78	99.76	0.02	35	96.61	96.15	0.46
14	99.76	99.74	0.02	36	96.15	96.00	0.15
15	99.74	99.70	0.04	37	96.00	93.53	2.47
16	99.70	99.56	0.14	38	93.53	92.98	0.55
17	99.56	99.54	0.02	39	92.98	90.97	2.01
18	99.54	99.53	0.01	40	90.97	84.54	6.43
19	99.53	99.48	0.05	41	84.54	51.57	33.03
20	99.48	99.37	0.11	42	51.51	0.00	51.51
21	99.37	99.29	0.08	43	0.00		
22	99.29	99.28	0.01				

D = Difference between adjacent similarity levels S_L and S_{L-1}, $_{SL}$ = Similarity level S_{L-1} = Similarity level before the level under consideration.

Table 4. Final partition of the 44 CMD-resistant varieties.

Cluster	No of observations	Within cluster sum of squares	Average distance from controid	Maximum distance from controid
1	3	1.830	0.728	0.977
2	15	32.102	1.336	3.039
3	23	62.971	1.568	2.539
4	2	0.581	0.539	0.539
5	1	0.000	0.000	0.000

CMD varieties 98/2226- 95/0379 (Figure 3) can also be said to be average with respect to moisture (X_1) and dry matter (X_7). The dendrogram of the cluster has 23 varieties including 21, 23, 27, 31, 38, 40, 32, 34, 19, 22, 25, 20, 24, 26, 28, 29, 30, 33, 36, 35, 37, 39 and 41.

Cluster 4 with CMD varieties 92B/0061 and 98/0581 (varieties 42 and 43) is moderately high in moisture (X_1) average in protein (X_2), low in ash (X_3), average in fat (X_4) and fibre (X_5) but low in carbohydrate (X_6) and dry matter (X_7) contents (Figure 4). Cluster 5 with only CMD variety 96/1569 (variety 44) is very high in moisture (X_1), average in protein (X_2), moderately high in ash (X_3), low in fat (X_4) and fibre (X_5), moderately high in carbohydrate (X_6) and least in dry matter (X_7) (Figure 5). The dendrograms of the data are shown as cluster groups in Figures 1-5. The dendrograms show the pattern of clustering among the varieties with connecting lines further to the right indicating more distance between varieties and clusters. The dendrogram is a hierarchical tree diagram or plot which shows the relative size of the proximity coefficient at which cases were combined. Cases with low distances are close together with a line linking them. A short distance from the left of the dendrogram indicates that they are agglomerated into a cluster at a distance coefficient indicating similarity.

Figure 1 shows the dendogram of first cluster, which includes varieties 1, 2 and 3. Figure 2 shows the dendogram of the second cluster which includes

Table 5. Sorted group ID for the cassava varieties.

S/N	CMD varieties	Group ID	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	97/4769	1	5.52	2.42	0.64	0.41	0.07	90.91	94.48
2	99/6012	1	6.17	1.93	0.64	0.24	0.06	91.38	93.83
3	94/0561	1	6.51	1.40	0.22	0.51	0.07	91.19	93.49
4	97/0162	2	6.97	2.07	0.32	0.58	0.05	90.06	93.03
5	94/0026	2	6.99	1.40	0.24	0.56	0.04	90.71	93.01
6	96/1642	2	7.02	2.55	0.39	0.04	0.04	90.60	92.98
7	98/0510	2	7.26	1.82	0.07	0.27	0.05	90.53	92.74
8	98/0505	2	7.29	2.45	0.05	0.20	0.06	89.95	92.71
9	99/3037	2	7.45	2.15	0.59	0.42	0.01	87.61	92.55
10	98/2101	2	7.46	1.93	0.53	0.42	0.06	87.37	92.54
11	97/4763	2	7.48	2.41	0.37	0.23	0.05	89.46	92.52
12	99/26324	2	7.53	1.40	0.31	0.44	0.06	89.62	92.47
13	97/2205	2	8.19	1.40	0.34	0.40	0.05	89.62	91.81
14	98/1565	2	8.19	0.70	0.51	0.34	0.07	89.50	91.81
15	92/0325	2	8.22	1.75	0.10	0.12	0.01	89.50	91.78
16	92B/00068	2	8.32	1.23	0.20	0.66	0.03	88.67	91.68
17	TME419	2	8.32	0.70	0.41	0.39	0.06	90.12	91.68
18	96/0603	2	8.37	0.98	0.22	0.39	0.08	89.96	91.63
19	98/2226	3	8.62	2.80	0.45	0.34	0.02	87.76	91.38
20	92/0058	3	8.70	1.78	1.58	0.34	0.02	87.58	91.30
21	97/0211	3	8.70	2.80	1.53	0.56	0.04	86.19	91.30
22	95/0289	3	9.09	1.78	0.22	0.42	0.02	88.47	90.91
23	92/0326	3	9.13	2.80	0.34	0.56	0.08	87.09	90.87
24	92/1452	3	9.14	1.40	1.28	0.64	0.02	87.52	90.86
25	98/0002	3	9.32	1.78	0.43	0.41	0.04	88.01	90.68
26	97/4779	3	9.53	1.10	0.97	0.42	0.04	87.12	90.65
27	96/1642	3	9.35	2.80	0.48	0.56	0.01	86.80	90.65
28	96/0523	3	9.43	1.21	0.44	0.17	0.04	89.71	90.57
29	M98/0068	3	9.53	1.05	0.20	0.17	0.04	89.01	90.47
30	M98/0028	3	9.55	0.35	0.37	0.44	0.03	89.26	90.45
31	96/1089A	3	9.97	2.80	0.94	0.43	0.06	86.21	90.43
32	95/0166	3	9.63	2.55	2.13	0.03	0.03	86.18	90.37
33	92/0057	3	9.66	1.05	0.57	0.43	0.01	88.28	90.38
34	96/1314	3	9.78	2.10	3.34	0.29	0.05	86.93	90.22
35	97/3200	3	9.78	1.40	0.17	0.36	0.07	88.22	90.22
36	98/0040	3	9.96	1.05	0.53	0.42	0.05	87.99	90.04
37	TM/530572	3	10.13	1.75	0.77	0.33	0.03	87.32	89.87
38	99/2123	3	10.26	2.80	0.42	0.53	0.08	85.91	89.74
39	92/0067	3	10.38	1.08	0.77	0.60	0.04	87.16	89.62
40	97/0039	3	10.39	2.10	0.52	0.50	0.06	86.06	89.61
41	95/0379	3	10.45	1.23	0.47	0.43	0.04	87.83	89.55
42	92B/0061	4	10.65	1.23	0.08	0.48	0.08	76.89	89.35
43	98/0581	4	11.26	1.26	0.53	0.02	0.04	76.89	88.74
44	96/1569	5	12.25	2.10	0.52	0.29	0.03	84.81	87.75

varieties 9, 10, 4, 6, 5, 7, 8, 11, 12, 13, 15, 14, 17, 18 and 16. Figure 3 shows the dendogram of the third cluster which has 23 varieties including varieties 21, 23, 27, 31, 38, 40, 32, 34, 19, 22, 25, 20, 24, 26, 28, 29, 30, 33, 36, 35, 37, 39 and 41. Figure 4 is the dendogram of the fourth cluster which are varieties 42 and 43. The last dendogram that contains only one variety namely variety 44 is shown in Figure 5.

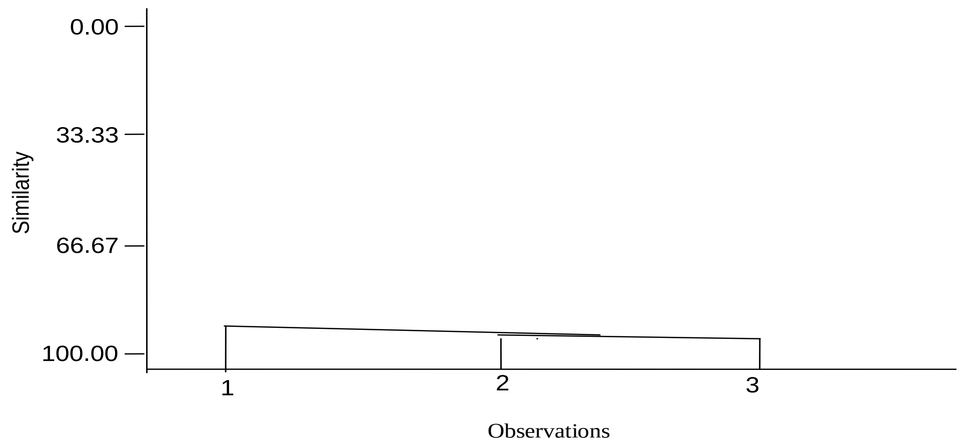

Figure 1. Dendogram for Cluster 1 of the 44 CMD-resistant varieties based on their proximate composition. The 3 observations (1, 2 and 3) are the cassava varieties (97/4769, 99/6012 and 94/0561) in the same cluster group based on similarities of their proximate composition.

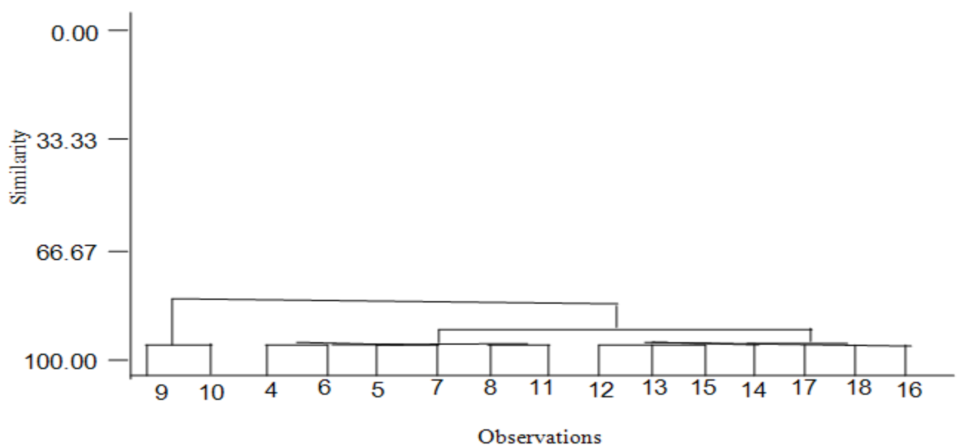

Figure 2. Dendogram for Cluster 2 of the 44 CMD-resistant varieties based on their proximate composition. The 15 observations (4-18) are the cassava varieties (97/0162- 96/0603) in the same cluster group based on similarities of their proximate composition.

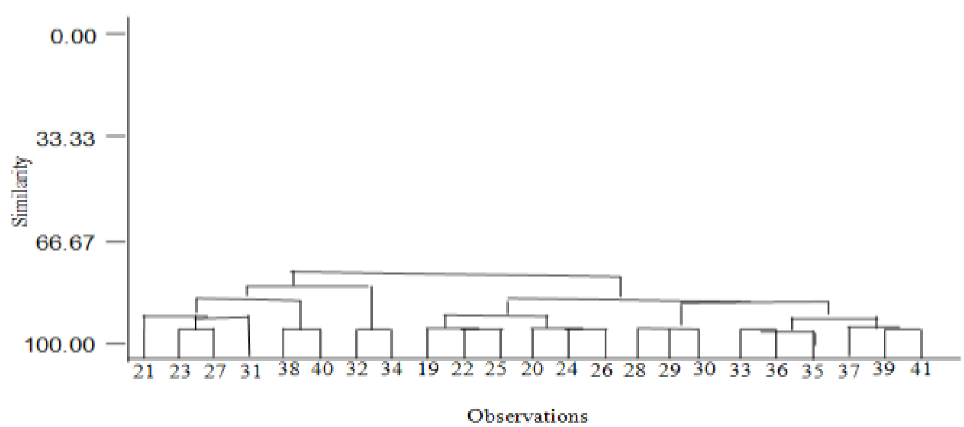

Figure 3. Dendogram for Cluster 3 of the 44 CMD-resistant varieties based on their proximate composition. The 23 observations (19-41) are the cassava varieties (98/2226-95/0379) in the same cluster group based on similarities of their proximate composition.

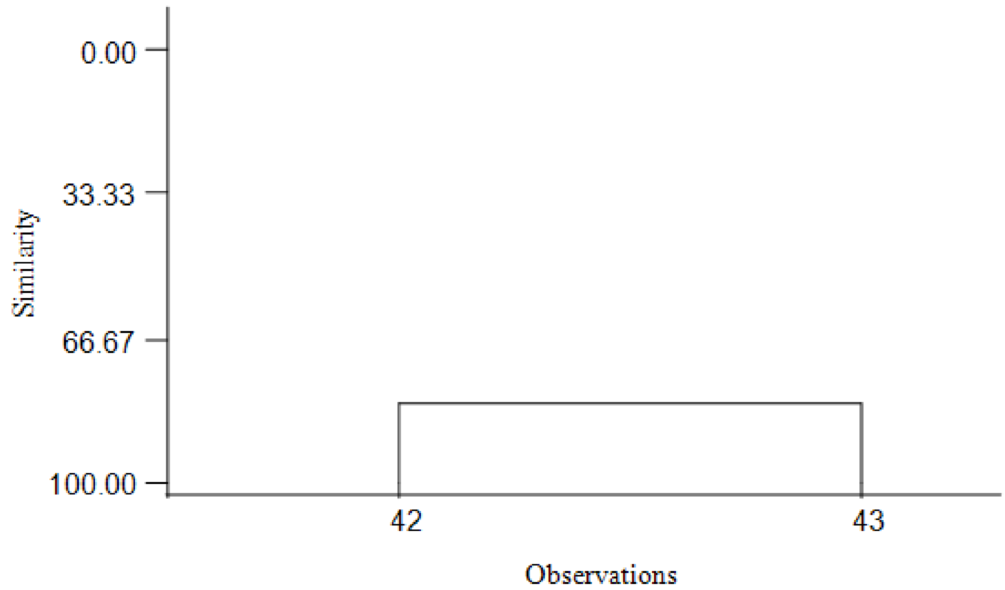

Figure 4. Dendogram for Cluster 4 of the 44 CMD-resistant varieties based on their proximate composition. The 2 observations (42 and 43) are the cassava varieties (92B/0061 and 98/0581) in the same cluster group based on similarities of their proximate composition.

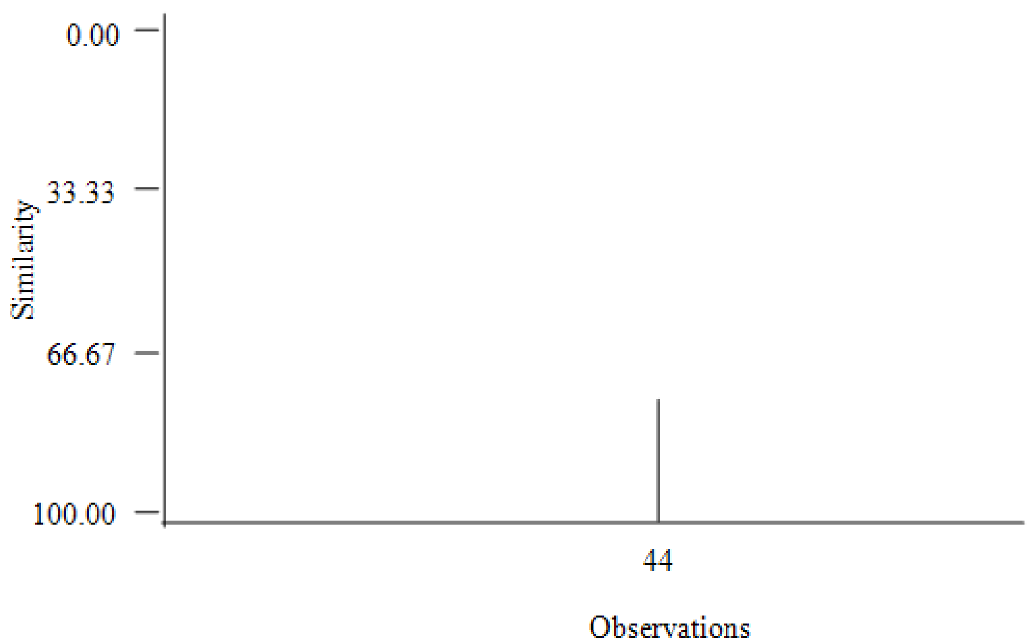

Figure 5. Dendogram for Cluster 5 of the 44 CMD-resistant varieties based on their proximate composition. The 44th CMD-resistant variety (96/1569) in the cluster group had similarity value different from the other 43 varieties.

Conclusion

Five distinct grouping were made out of the 44 different cassava varieties based on their similarity level with respect to their *fufu* flour compositions. It is observed that varieties 1, 2 and 3 are in the first group. Varieties 4 to 18 are in the second group, cassava varieties 19 to 41 make up the 3rd group. Varieties 42 and 43 are in the fourth group and finally, cassava variety 44 is the only variety in the fifth group. The recommendation from this study is that since the study has succeeded in grouping the 44 varieties into five groups, the farmers need to grow only five out of the 44 varieties, one from each group and have almost all the benefit of growing all the 44 varieties at a time.

REFERENCES

Anderberg MR (1973). Cluster Analysis for applications. Academic Press, Inc., New York.

Crawford IM, Lomas RA (1980). Factory Analysis - A Tool for Data Reduction. Eur. J. Mark. 14(7):414-421. DOI: 10.1108/EUM0000000004917. Emerald Group Publishing Limited.

Cunningham KM, Ogilvie JC (1972). Evaluation of hierarchical grouping techniques: A preliminary study. Comput. J. 15(3):209-213.

Diday E, Simon JC (1976). Clustering analysis. In Digital Pattern Recognition, K. S. Fu, Ed. Springer-Verlag, Secaucus, NJ. pp. 47-94.

Etudaiye HA, Nwabueze TU, Sanni LO (2009). Quality of *fufu* processed from Cassava Mosaic Disease (CMD) resistant varieties. Afr. J. Food Sci. 3(3):061-067.

Everitt BS (1993). Cluster Analysis. Edward Arnold, Ltd., London. UK.

FAO (2002). Food and Agriculture Organization of the United Nations, Agricultural towars 2015/30. Technical interim Report, April, 2000. Rome.

FAO (2004). Food and Agriculture Organization of the United Nations, Online statistical database. Rome, Italy, April 1, 2004. www.fao.org.

Friedman HP, Rubin J (1967). On some invariant criteria for grouping data. J. Am. Statist. Assoc. 62:1159-1178.

IITA (2005). Growing cassava commercially in Nigeria. Cassava illustration guide book. International Institute of Tropical Agriculture, Ibadan, Nigeria. pp. 21-22.

Jain AK, Dubes RC (1988). Algorithms for Clustering Data. Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River. NJ.

Jain AK, Murty MN, Flynn PJ (1999). Data clustering: A review. ACM Comput. Surveys 31(3):264-323.

Michalski R, Stepp RE (1983). Automated construction of classifications: conceptual clustering versus numerical taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-5. 5:396-409.