MEDWELL PUBLISHING Journal of Modern Mathematics and Statistics 3 (1): 22-24, 2009
ISSN: 1994-5388

Principal Component Analysis of Nutritional Quality of 43 Cassava Varieties

Nwabueze Joy Chioma and Ereh Trinitas
Department of Statistics, Abia State University, Uturu, Abia State, Nigeria

Abstract: Data on nutritional quality of fufu flour produced from 43 cassava varieties were analyzed using multivariate methods. The cassava varieties were selected from the newly developed Cassava Mosaic Disease resistant varieties in Onne, Nigeria. Analysis showed that the 1st Principal Component (PC_1) explains about 26.37%, the 2nd Principal Component explained about 21.90 (PC_2) the 3rd Components (PC_3) explained about 15.43% and the 4th Principal Component (PC_4) explained about 14.77%. The 1st, 4 principal components accounted for about 78% of the total variation.

Key words: Multivariate method, principal component, cassava varieties, proximate composition, eigen values, population parameters, correlation matrix

INTRODUCTION

Agriculture plays an important role in the economic development of Nigeria. It provides food and employment for the population, raw materials and foreign exchange earning for the development of industrial sector. Cassava cultivation increased after 1850 in the East African Territories as a result of the efforts of Europeans and Arabs who were pushing into the interior and recognized cassavas as a safe guard against the frequent period of famine.

Nigeria is one of the largest producers of cassava in the world. Its production is currently put at about 34 mm tones a year. In addition to cassava, being produced primarily for food in form of garri, lafun and fufu (Sanni *et al*., 2006), it is also processed into several secondary products of industrial market value, which include chips, pellets, flour, adhesives, alcohol and starch. These products, which constitute vital raw materials for livestock feed, alcohol/ethanol, textile, confectionary, wood, food and soft drinks industries are traceable in the international market (FAO, 2004).

Cassava verities recently developed for pest and 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).

This research analyses a set of data on nutritional composition of fufu flour produced from 43 CMD resistant varieties of cassava planted by the in International Institute of Tropical Agriculture (IITA, 2005) at Onne, Port Harcourt, Nigeria (Etudaiye *et al*., 2008). The nutritional composition of the fufu flours measured on these varieties included moisture content, protein, ash, fat, fiber, carbohydrate and dry matter. This set of data was analyzed using a multivariate analysis method called principal component. Principal component analysis is a multivariate analysis that involves data reduction and data interpretation. Ian (2005) stated that when large multivariate datasets are analyzed, it is often desirable to reduce their dimensionality using Principal component analysis technique. It replaces the original variables by a smaller number of derived variables, the principal components, which are linear combinations of the original variables. Often, it is possible to retain most of the variability in the original variables with the smaller number very much smaller than original variables.

A number of researchers had used principal components in data analysis including Muluneh *et al.* (2008) and Adebowale and Onitilo (2008) that carried out a research on composition and functional properties in germ plasma for diversity and potential of yam for food and non-food applications. Their findings among others showed that starch content varied from 65.2-76.6% dry matter, while the protein content range was between 6.47 and 30.6%. Adebowale and Onitilo (2008) used principal component analysis on the chemical composition of tapioca grits from different cassava varieties and roasting methods. Their results showed that moisture, starch and sugar content accounted for 83% of the variations in the chemical composition of the tapioca samples. The objective of this study, was to locate the nutritional composition that contributed maximum variability of the fufu flour processed from the 43 CMD cassava varieties and detect the structure of the data and

Corresponding Author: Nwabueze Joy Chioma, Department of Statistics, Abia State University, Uturu, Abia State, Nigeria

reduce its number. The study was also to examine the percentage contribution of each composition to the total variation in order to assess the variance.

MATERIALS AND METHODS

Theoretical framework: The principal components can be estimated using the population parameters.

$$\left.\begin{array}{lll} y_1 & = & L_{11}X_1 + \cdots + Lp_1Xp \\ y_2 & = & L_{12}X_1 + \cdots + Lp_2Xp \\ \vdots & & \vdots \\ y_p & = & L_1pX_1 + \cdots + LpXp \end{array}\right\} \qquad (1)$$

where:

$y_1, y_2, \dots y_p$ = The principal components, which are the linear combinations of the original

p = Components variables, which in this study are the proximate compositions

$$\text{var}\,(y_1) = L\Sigma L$$

where:

Σ = The correlation matrix of the x variables. If y_i and y_i are members of y then

$$\text{cor}\,(y_i\ y_{i'}) = 0 \text{ for all } i \neq i'$$

The constants L_{i1}, L_{i2}, ... L_ip are the elements of the corresponding eigenvectors normalized so that $L'_i\ L_i = 1$. Thus, the 1st principal component is the vector L_i, which maximizes var (L'_iX) and is of length 1. Similarly, the ith principal component is that vector L_i, which maximizes $(L_iX\ )$ subject to $L'_i\ L_i = 1$.

The principal component could also be expressed in terms of the eigen value of the original variable x hence, the ith principal component is given by

$$y_i = e'_iX \quad i = 1, 2, \dots, \rho \qquad (2)$$

where:

e_i = The eigen vector of Σ, the covariance matrix of the compositions

$(\lambda_1\ e_1)$ = The eigen value-eigen vector pair of Σ and $(\lambda_1 \geq \lambda_2 \geq \dots \lambda\rho \geq 0)$

RESULTS AND DISCUSSION

Mean vector and correlation matrix: The compositions constitute the variables for this study,

where:

x_1 = Moisture
x_2 = Protein
x_3 = Ash
x_4 = Fat
x_5 = Fiber
x_6 = Carbohydrate
x_7 = Dry matter

The correlation between the variables was calculated and the result is displayed on Table 1. It shows that all the variables except x_4 are negatively correlated with moisture content x_1. Protein is positively correlated with to fat and dry matter while, it is negatively correlated with ash, fiber and carbohydrate. In summary all the variables are correlated with each other some negatively and some positively though not very highly. The eigen values of the variables were also calculated and the result is shown in Table 2. Moisture content (x_1), protein (x_2), ash (x_3), fat (x_4), fiber (x_5) carbohydrate (x_6) and dry matter (x_7). The 1st column of Table 2 shows the variables used for this study. The 2nd column of Table 2 is the eigen values of the variables, which are 1.84591, 1.53304, 1.08011, 1.03419, 0.81937, 0.50430 and 0.1830, which sum to 7.0000, which is the total number of variables. Since, the correlation matrix is used, the total variance to be partitioned between the components is equal to the number of variables. The 3rd column gives the proportion of variation associated with each variable, which is the ration of the eigen value of variables to the total variables. The last column showed the cumulative proportion of variation, up to each variables.

Four principal components were retained for this study because there are only 4 components, whose eigen values are >1 (Kaiser, 1960). The eigenvectors are termed component scores because they give scores to the principal components for example; the 1st principal component, which is the most important component has a score of

$$y_1 = 0.28496x_1 + 0.45267x_2 + (-)0.10376x_3 + 0.43487x_4 + (-)$$
$$0.31807x_5 + (-)0.64136x_6 + 0.03850x_7$$

where, the coefficient (0.28496, 0.45267,..., 0.03858) are the normalized eigenvectors.

From Table 3, the 1st principal component explains 26.37% of the total variation in this study. The other principal component scores are calculated in the same way. The 2nd-4th principal components explain, respectively, 21.90, 15.43 and 14.77% of the total variation in the study. The other component PC_5, PC_6 and PC_7 contribute very little to the total variation and we recommend that they be neglected.

Selection of important variables in 1st 4 principal components: The variables that have scoring coefficient

Table 1: Correlation coefficients of the variables

Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1.0000						
X_2	-0.2503	1.0000					
X_3	-0.0391	-0.0796	1.0000				
X_4	0.1980	0.0846	0.0585	1.0000			
X_5	-0.1808	-0.0880	0.0992	-0.1840	1.0000		
X_6	-0.3360	-0.5839	-0.0262	-0.3127	0.1524	1.0000	
X_7	-0.2272	0.1745	-0.0106	0.2070	0.2668	-0.0189	1.0000

Table 2: Eigen values and scoring coefficient of the variables

Variables	Eigen value	Proportion	Cumulative
X_1	1.84591	0.2637	0.2637
X_2	1.53304	0.2190	0.4827
X_3	1.08011	0.1543	0.6370
X_4	1.03419	0.1477	0.7847
X_5	0.81937	0.1170	0.9018
X_6	0.50430	0.0720	0.9738
X_7	0.18308	0.0262	1.0000

Table 3: Component loadings of the 7 variables

Variables	PC_1	PC_2	PC_3	PC_4
1	0.28496	-0.54669	0.36121	0.10305
2	0.45267	0.46121	-0.44646	0.09456
3	-0.10376	0.03423	0.19439	0.87958
4	0.43487	0.00351	0.54747	-0.22486
5	-0.31807	0.38266	0.32780	0.19039
6	-0.64136	-0.07549	0.05090	-0.28824
7	0.03858	0.057889	0.47196	-0.19208

of at least 50% in absolute terms are retained and selected in the 4 principal component of the variables retained from the 1st principal component only carbohydrate x_6 is selected. Moisture content x_1 and dry matter x_7 were selected in the second principal component. In the 3rd and 4th principal components, fat (x_4) and variable (x_3), ash was selected, respectively.

CONCLUSION

The result of the principal component analysis of the data showed that the variables can be presented adequately in just 4 dimensions because, we obtained only 4 principal components. The 1st principal component is associated with protein and carbohydrate because they have high loadings. The 2nd principal component is associated with moisture and dry matter. The 3rd and 4th principal components are associated with fat and ash, respectively. These variables explain up to 78% of the total variance. The percentage contributions of each of the 4 principal components to the total variations are 26.37, 21.90, 15.43 and 14.77%, respectively.

The correlation matrix showed that there is a significant negative relationship between carbohydrate (x_6) and protein (x_2) on one hand and between carbohydrate and moisture at 5% level of significance.

REFERENCES

Adebowale, S.L.O. and M.O. Onitilo, 2008. Chemical composition and pasting properties of Tapioca grits from different Cassava varieties and roasting methods. Afr. J. Food Sci., 2: 77-82. http://www.academicjournals.org/ajfs.

Etudaiye, H.A.I., T.U. Nwabueze and L.O. Sanni, 2008. Quality of fufu processed from Cassava Mosaic Disease (CMD) resistant varieties. AJFS, Vol. 2 (in Press). http://www.academicjournals.org/ajfs.

FAO, 2004. Online Statistical Database Rome, Italy; Food and Agriculture Organization of the United Nations. www.fao.org.

Ian, J., 2005. Principal Component Analysis. University of Aberdeen, Aberdeen, UK. Pub. By John Wiley and Sons, Ltd. DOI: 10.1002/0470013192.bsa501.

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

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. EPM, 20: 141-151.

Muluneh, Maass, Tamru, Brigittle, Elike and Pawaltik, 2008. Composition and functional properties in germplasm for diversity and potential of yam for food and non-food application. JSFA, 88 (10): 1675-1685.

Sanni, O.L., A.A. Adebeowale, T.A. Filani, O.B. Oyewole and A. Westby, 2006. Quality of flash and rotary dried fufu flour. JFAE, 4 (3x).