

APPLICATION OF MULTIVARIATE STATISTICAL  
METHOD IN THE STUDY OF MORPHOLOGICAL  
FEATURES OF TILAPIA CABREA

By

Bartholomew A. Uchendu (Ph.D)

Department of Maths/Statistics, Federal Polytechnic, Nekede,  
Owerri, Nigeria

# Abstract

Data was collected on the morphological features of Tilapia Cabrea. The weight and lengths were measured in grammes and millimeters respectively. The data was subjected to a multivariate data analysis; the principal component analysis and the analysis showed that four principal components accounted for about 88% of the total variability viz: body weights ( $X_3$ ), body depth ( $X_8$ ), Snout length ( $X_6$ ) and standard length ( $X_2$ ). The analysis also included the total length ( $X_1$ ) as one of the least contributors to the size of the fish. This could be justified as it known that the fins and distance between the anterior and posterior extremity of the mouth of the fish are like chaff and have no weights.

**KEY WORDS:** Principal Components, Loadings, Variance – Covariance, Transformation, K-variables

# 1.0 INTRODUCTION

In a multivariate observation different variables make their individual contributions and our interest now lies in investigating the relationship of these different variables and to determine the extent of variation between them.

Given the complexities of such situation with regard to the number of interrelated variables that have to be considered, King (1969) suggests that it is not surprising that multivariate procedures and techniques will be favoured in the solution of such problems.

It is observed that the individual contributors of these different variables make up the size, quality or value of the object and hence by studying their interrelationships, we can extract or identify the meaningful variables, and therefore do away with the less important variables. With the identification of the important variables meaningful interpretations on the variables are hence achieved.

The principal component analysis is a variable reducing scheme and is applied in different areas and fields of which some are in agriculture, crime, geography, education and forestry to mention but a few.

We shall adopt this particular method to analyse a fish (Tilapia Cabrea) in order to project the importance of multivariate analysis in the solution of our diverse problems. The techniques of principal component analysis was generalized and the solution outlined originally by Harold Hotelling. See King (1969), Press (1972) and Cooley and Lohnes (1971). The method is also described as a process of eliminating variables, which contribute relatively little extra information.

In the principal component, it is observed that each member bears the values of  $P$  variates, say,  $x_1, \dots, x_P$ , it is generally possible to find a linear transformation to  $P$  new variates which are independent and also account in turn for as much of the variation as possible in the sense that the variation of the first component is the maximum among all linearly transformed variates. The variance of the second component is maximum among all linearly transformed variates orthogonal to the first component and so on. See Kendal and Buckland (1975).

The results hence obtained from these sets of variables are interpreted thus: five or less components are retained or extracted which has to account for about 75% or more of the variance. These extracted components accounting for the above percentages range are in most cases used for the interpretations. See Stevens, (1976).

# 1.1 NOTATIONS AND DEFINITION OF TERMS

- $X_1$  = Total Length (TL) is the distance between the anterior extremity of the mouth to the posterior extremity of the caudal fin.
- $X_2$  = Standard Length (SL) is the distance between the posterior extremity of the mouth to the posterior end of the caudal peduncle.
- $X_3$  = Body weight (BW) is the total body weight of the fish.
- $X_4$  = Gutted Weight (GW) is the total weight of the fish minus the weight of the gut.
- $X_5$  = Head Length (HL) is the distance between the anterior extremity of the mouth to the posterior end of the gill cover.
- $X_6$  = Snout Length (NL) is the distance between the anterior extremity of the mouth to the nasal opening.
- $X_7$  = Ocular Length (OL) is the diameter of the eye
- $X_8$  = Body Depth (BD) is the longest distance between the ventral and dorsal surface of the fish.
- $X_9$  = Head Width (HW) is the distance from one side of the head to another (cheek to cheek)
- $X_{10}$  = Mouth Length (ML) is the distance from one end of the mouth to the other.

All measurements of  $X_1$ ,  $X_2$ ,  $X_5$ ,  $X_6$ ,  $X_7$ ,  $X_8$ ,  $X_9$ , and  $X_{10}$  are in millimeters and  $X_3$  and  $X_4$  are in grammes.

# 2.0 METHODOLOGY

The solution of the principal component analysis was originally outlined by Hotelling. The basic technique of this multivariate method is well described by Pielou (1969), King (1969), Cooley and Lohnes (1971) and Koutsoyiannis (1977) and many others. These authors had their different approaches but which boiled down to the same results.

The aim of the principle components is to construct out a set of variables,  $X_j$ 's  $j = 1, 2, \dots, k$ , of new variables  $P_i$  called principal components, which are linear combinations of the  $X$ 's:

$$\left. \begin{aligned} P_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1k}X_k \\ P_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2k}X_k \\ &\cdot \\ &\cdot \\ &\cdot \\ &\cdot \\ P_k &= a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kk}X_k \end{aligned} \right\} 2.1$$

The method of principal components could be applied using original values of the  $X_j$ 's or deviations from their  $X_j = X_j - \bar{X}_j$ , or the standardized variables (measured as the deviations of the  $X_j$ 's from the means and subsequently divided by the standard deviations).

$$Z_j = \frac{X_j}{S_{xj}} \quad 2.2$$



We will adopt this later procedure of standardized variables, because it is more general in the sense that it can be applied when variables are measured in different unit.

This transformation by standardization makes all the variables dimensionless and hence follows a normal distribution with zero mean and unit variable

$$Z \sim N(0, 1)$$

It is noted however that the values of the principal components will be different depending on the way in which the variables are used - whether original, their deviations or their standardized values.

The  $a$ 's called loadings, are chosen so that the constructed principal components satisfy two conditions:

- I. The principal components are uncorrelated (orthogonal)
- II. The first principle component  $P_1$  absorbs and accounts for the maximum possible proportion of the total variation in the set of all  $X$ 's, the second principal component  $P_2$  absorbs and accounts for the maximum of the remaining variation in the  $X$ 's after showing for the variation accounted for by the first principal components, and so on.

King (1969) suggested using standardized variables to obtain the variance - covariance matrix. The variance - covariance matrix becomes the matrix of simple correlations  $R$ .

Thus

$$\mathbf{X}'\mathbf{X} = \mathbf{R}$$

Hence the matrix

$$\begin{array}{ccc}
 \frac{\sum X_1^2}{\sqrt{\sum X_1^2}} & \frac{\sum X_1 \sum X_2}{\sqrt{\sum X_1^2} \sqrt{\sum X_2^2}} & \dots \dots \dots \frac{\sum X_1 X_k}{\sqrt{\sum X_1^2} \sqrt{\sum X_k^2}} \\
 \frac{\sum X_1 \sum X_2}{\sqrt{\sum X_1^2} \sqrt{\sum X_2^2}} & \frac{\sum X_2}{\sqrt{\sum X_2^2}} & \dots \dots \dots \frac{\sum X_2 \sum X_k}{\sqrt{\sum X_2^2} \sqrt{\sum X_k^2}} \\
 \cdot & \cdot & \cdot \\
 \cdot & \cdot & \dots \dots \dots \cdot \\
 \cdot & \cdot & \cdot \\
 \frac{\sum X_1 \sum X_k}{\sqrt{\sum X_1^2} \sqrt{\sum X_k^2}} & \frac{\sum X_2}{\sqrt{\sum X_2^2} \sqrt{\sum X_k^2}} & \dots \dots \dots \frac{\sum X_k}{\sqrt{\sum X_k^2}}
 \end{array}$$

The main diagonals have unit values since they are self correlated i.e.

$$(r_{xixi} = 1, \forall_i S)$$



79 observations on 10 variables (morphological features of Tilapia Cabrea) were subjected to a principal component analysis.

Since the observations from the data are in different units, that is  $X_3$ ,  $X_4$  are measured in grammes and the rest in millimeters, we standardize the data using the form

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j} \quad 2.4$$

Where  $i = 1, 2, \dots, 79$ ,  $j = 1, 2, \dots, 10$

$X_{ij}$  are observations

$\bar{X}_j$  are the means of the variables

$S_j$  are the standard deviations of the variables

This standardization makes our observation dimensionless

Table 2.1: Statistic of the data

|          | <b>Mean (<math>\bar{X}_j</math>)</b> | <b>Standard deviations (<math>S_j</math>)</b> |
|----------|--------------------------------------|---|
| $X_1$    | 118.3417722                          | 19.863542                                     |
| $X_2$    | 92.97468354                          | 15.52167606                                   |
| $X_3$    | 45.51835443                          | 16.72884308                                   |
| $X_4$    | 39.18632911                          | 15.42382823                                   |
| $X_5$    | 32.31645570                          | 4.91085742                                    |
| $X_6$    | 12.36708861                          | 2.62716194                                    |
| $X_7$    | 10.21518987                          | 1.59060451                                    |
| $X_8$    | 45.74683544                          | 6.98252095                                    |
| $X_9$    | 17.81012658                          | 3.09290316                                    |
| $X_{10}$ | 17.39240506                          | 2.91438342                                    |

Using equation 2.4

$$Z_{1,1} = \frac{135 - 118.3417722}{19.8635642} = 0.8386$$

.

.

.

$$Z_{1,10} = \frac{18 - 17.39240506}{2.91438342} = 0.2066$$

.

.

.

$$Z_{79,10} = \frac{20 - 17.39240506}{2.91438342} = 0.8865$$

The standardized values are hence obtained and will be in the form as presented in table 2.2

Table 2.2 Standardized Values

|           |                |                |                |                |                |                |                |                |                |                |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| <b>1</b>  | <b>0.8386</b>  | <b>-0.5138</b> | <b>-0.2271</b> | <b>-0.0348</b> | <b>-0.2681</b> | <b>-0.5206</b> | <b>-0.1353</b> | <b>-0.3934</b> | <b>-0.5853</b> | <b>0.2066</b>  |
| <b>2</b>  | <b>-0.1179</b> | <b>-3.3849</b> | <b>-0.3777</b> | <b>-0.0892</b> | <b>-0.4717</b> | <b>-0.901</b>  | <b>-0.4934</b> | <b>0.6091</b>  | <b>0.3847</b>  | <b>-2.5133</b> |
| .         | .              | .              | .              | .              | .              | .              | .              | .              | .              | .              |
| .         | .              | .              | .              | .              | .              | .              | .              | .              | .              | .              |
| .         | .              | .              | .              | .              | .              | .              | .              | .              | .              | .              |
| <b>79</b> | <b>-0.1179</b> | <b>-0.1917</b> | <b>-0.005</b>  | <b>-0.058</b>  | <b>-0.064</b>  | <b>0.5204</b>  | <b>-0.135</b>  | <b>-0.11</b>   | <b>-0.262</b>  | <b>-0.8865</b> |

The matrix of intercorrelation is given in table 2.3

|                       | <b>X<sub>1</sub></b> | <b>X<sub>2</sub></b> | <b>X<sub>3</sub></b> | <b>X<sub>4</sub></b> | <b>X<sub>5</sub></b> | <b>X<sub>6</sub></b> | <b>X<sub>7</sub></b> | <b>X<sub>8</sub></b> | <b>X<sub>9</sub></b> | <b>X<sub>10</sub></b> |
|-----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-----------------------|
| <b>X<sub>1</sub></b>  | <b>1.000</b>         | <b>0.6893</b>        | <b>0.7267</b>        | <b>0.7333</b>        | <b>0.5655</b>        | <b>0.4752</b>        | <b>0.4359</b>        | <b>0.5729</b>        | <b>0.5326</b>        | <b>0.4328</b>         |
| <b>X<sub>2</sub></b>  | <b>0.6893</b>        | <b>1.000</b>         | <b>0.8007</b>        | <b>0.7488</b>        | <b>0.7851</b>        | <b>0.7111</b>        | <b>0.5257</b>        | <b>0.7041</b>        | <b>0.6870</b>        | <b>0.5534</b>         |
| <b>X<sub>3</sub></b>  | <b>0.7267</b>        | <b>0.8007</b>        | <b>1.000</b>         | <b>0.9342</b>        | <b>0.7939</b>        | <b>0.7699</b>        | <b>0.5182</b>        | <b>0.7740</b>        | <b>0.6587</b>        | <b>0.5087</b>         |
| <b>X<sub>4</sub></b>  | <b>0.7333</b>        | <b>0.7488</b>        | <b>0.9342</b>        | <b>1.000</b>         | <b>0.7161</b>        | <b>0.6604</b>        | <b>0.4873</b>        | <b>0.7068</b>        | <b>0.6580</b>        | <b>0.4758</b>         |
| <b>X<sub>5</sub></b>  | <b>0.7655</b>        | <b>0.7851</b>        | <b>0.7939</b>        | <b>0.7161</b>        | <b>1.000</b>         | <b>0.8117</b>        | <b>0.4983</b>        | <b>0.7206</b>        | <b>0.6328</b>        | <b>0.5531</b>         |
| <b>X<sub>6</sub></b>  | <b>0.4752</b>        | <b>0.7111</b>        | <b>0.7699</b>        | <b>0.6604</b>        | <b>0.8117</b>        | <b>1.000</b>         | <b>0.6497</b>        | <b>0.8270</b>        | <b>0.6635</b>        | <b>0.6414</b>         |
| <b>X<sub>7</sub></b>  | <b>0.4359</b>        | <b>0.5257</b>        | <b>0.5182</b>        | <b>0.4873</b>        | <b>0.4783</b>        | <b>0.6497</b>        | <b>1.000</b>         | <b>0.6629</b>        | <b>0.6130</b>        | <b>0.6366</b>         |
| <b>X<sub>8</sub></b>  | <b>0.5729</b>        | <b>0.7041</b>        | <b>0.7740</b>        | <b>0.7068</b>        | <b>0.7206</b>        | <b>0.8270</b>        | <b>0.6629</b>        | <b>1.000</b>         | <b>7487</b>          | <b>0.6766</b>         |
| <b>X<sub>9</sub></b>  | <b>0.5326</b>        | <b>0.6870</b>        | <b>0.6587</b>        | <b>0.6580</b>        | <b>0.6318</b>        | <b>0.6635</b>        | <b>0.6130</b>        | <b>0.7487</b>        | <b>1.000</b>         | <b>0.6396</b>         |
| <b>X<sub>10</sub></b> | <b>0.4328</b>        | <b>0.5534</b>        | <b>0.5087</b>        | <b>0.4758</b>        | <b>0.5531</b>        | <b>0.6414</b>        | <b>0.6366</b>        | <b>0.6766</b>        | <b>0.6396</b>        | <b>1.000</b>          |
| <b>Σ</b>              | <b>6.1642</b>        | <b>7.2052</b>        | <b>7.4850</b>        | <b>7.1207</b>        | <b>7.0771</b>        | <b>7.2099</b>        | <b>6.0276</b>        | <b>7.3936</b>        | <b>6.8339</b>        | <b>6.1180</b>         |

The loadings  $l_{ij}$  are obtained using

$$l_{ij} = \frac{\sum_i^k \sum_j^k r_{xixj}}{\sqrt{\sum_i^k \sum_j^k r_{xixj}}} \quad 2.5$$

Hence for the first principal component  $P_1$ , our loading  $l_{ij}$ 's are obtained to be 0.7441, 0.8697, 0.9035, 0.8595, 0.8703, 0.7276, 0.8925, 0.8249 and 0.7387 respectively.

The first principal component is

$$P_1 = 0.7441Z_1 + 0.8697Z_2 + 0.9035Z_3 + 0.8595Z_4 + 0.8544Z_5 + 0.8703Z_6 + 0.7276Z_7 + 0.8925Z_8 + 0.8249Z_9 + 0.7387Z_{10}$$

# CONCLUSION

The principal component analysis has shown from the 79 observations on 10 variables (morphological features) on Tilapia Cabrea that there exists four major contributors to the size of the fish via body weight ( $X_3$ ), body depth ( $X_8$ ), snout length ( $X_6$ ) and standard length ( $X_2$ ) in order of magnitude.

The non-distinguishable features which are also the least contributors to the fish size, the oscular length ( $X_7$ ), mouth length ( $X_{10}$ ) and the total length ( $X_i$ ).

It is not surprising that the analysis included the total length ( $X_i$ ) as one of the least contributors to the fish size. This could be justified as it is known that the fins and the distance between the anterior and posterior extremity of the mouth of the fish are like chaff and have no weight. It is observed that the fins, and the distance between the anterior and posterior extremity of the mouth is the between the total length and the standard length of the fish.



# REFERENCES

- Ahmad, B. (1967), 'An Analysis of Crimes by the method of Principal Components' Applied Statistics, vol. 16 No. 1 pp 17 - 35
- Anderson, T. W (1963), Assyntotic Theory for Principal Components Analysis (Pp. 122 - 130)
- Cooley, W.W. and Lohnes, P.R. (1971) Multivariate Data Anaysis, John Wiley and Sons Inc. pp. 96-117.
- Jeffers, J.N.R (1967), "Two case studies in the "Application of the Principal component Analysis" Applied statistic vol 16, NO 3 pp 225236.
- Koutsoyiannis, A (1977), Theory of Econometrics 2<sup>nd</sup> Edition, Macmillan Press Ltd, London pp 424-434
- Pielou, E.C. (1969), An introduction to Mathematical Ecology, Wiley Interscience Press pp 283-301
- Press, S.J. (1972), Applied multivariate Analysis, Holt, Rinehart and Winston, Inc. pp283-301
- Stevens, J. (1986), Applied Multivariate Statistics for the Social Science, pp 337-339.
- Uchendu, B.A. (1998), Multivariate Analysis of Oil Palm (*Elaies Guineensis*) Data. IMSU PGD Project. Dept of Statistics IMSU pp. 138.