Statistical Analysis of Measurements Needed to Sew a Pair of Trousers

S. U. Enogwe[1]; D. F. Nwosu [2]; E. C. Ngome[3]; E. U. Oti[4]

[1]Department of Statistics,
Michael Okpara University of Agriculture, Umudike, Nigeria.

[2,3]Department of Mathematics/Statistics,
Federal Polytechnic Nekede, Owerri, Nigeria.

[4]Department of Statistics,
Federal Polytechnic, Ekowe, Bayelsa State, Nigeria.
E-mail**:** senogwe@yahoo.com[1]

***Abstract* — This paper demonstrates the use of the multivariate statistical technique of principal component analysis in determining the underlying structures of measurements taken by tailors to sew a pair of trousers. Interestingly, the measurements can be presented adequately in three dimensions because we obtained only three principal components. The first dimensions show that waist, length, lap, knee, and flap are significant for sewing a pair of trousers. The second dimension shows that four measurements (length, lap, flap, and base) could be taken together during trouser production and the third dimension suggests the use of four measurements (waist, length, knee, base) in sewing a pair of trousers.**

Keywords - *Zero-Inflated Poison Model, under-five children; spatial analysis; nonlinear effect; treated nets, South-West Nigeria.*

I. Introduction

In a contemporary society people are often judged based on the type of clothes they wear. It may be noted that one's appearance in clothes often creates a picture of his or her socio-economic status. Thus, the way garments fits one's body becomes an important factor in selection of clothes. It is obvious that a garment made of the best fabric cannot give the appearance of quality if it gives an individual poor fit. In fact, improperly fit garments do not appear attractive and cannot give an individual a feeling of physical comfort. Although, fit is a matter of choice, for instance, some people prefer ample ease garments while others feel more comfortable in body-fitted garments, it is necessary for a garment to fit an individual very well.

It is essential to note that humans come in varieties of shapes, sizes and proportions. The body characteristic is heavily relied on while making choices of clothes. Of course, there is no perfect body type that determines the size and style of clothes an individual can wear. Moreover, a good idea of one's shape may be determined by measuring the circumference of the waist, hip and chest. Height is also an important factor because two different people having the same height may entirely have different proportions. For instance, there are people with long arms and legs while others have long-waist. These body characteristics make a big difference in the kinds of style that fit an individual.

To achieve the ultimate goal of garment-fits, tailors take several body measurements before sewing clothes. These measurements generally are expected contribute to the outlook of the clothes. It worrisome that even after all these measurements have been taken, some of the constructed garments appear to fit individuals very poorly.

The aim of this paper is to analyze several measurements taken to sew a pair of trousers in order to determine whether the total variation in the measurements could be explained by a fewer constructs. The essence of such analysis is to guide tailors with the basic (principal) measurements to be taken during the course of trouser-making so as to reduce the variation that exist in the actual output of trousers. Most importantly, this work will benefit industrial trouser tailors who mass-produce trousers for people they do not have their actual measurements. Further, individuals who are opportune to read this paper would benefit from the work in the sense that it reveals to them the basic measurements they could give to tailors to sew trousers for them even when they are not physically available for such measurements. The paper is unfolded as follows: Section 2 deals with methodology. Section 3 presents results and discussion. The paper is concluded in Section 4.

II. METHODOLOGY

Six measurements trouser measurements (Waist, Length, Lap, Knee, Flap, Base) taken on 100 individuals were collected from 10 tailors at Psychiatric Tailoring Center, Aba. The data were analyzed using Principal Component Analysis (PCA).

2.1 The Principal Component Analysis (PCA)

According to Johnson and Wichern (2007), a Principal Component Analysis (PCA) starts with data on p correlated variables $\left(X_1, X_2, \ldots, X_p\right)$ on n individuals as shown in Table 1.

.

Table 1: The data layout for Principal Component Analysis (PCA)

Individual (i)	Variable Measurement			
	1	2	$\cdots$	p
1	X_{11}	X_{12}	$\cdots$	X_{1p}
2	X_{21}	X_{22}	$\cdots$	X_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	X_{p1}	X_{p2}	$\cdots$	X_{pp}

As orchestrated by Onyeagu (2003), since variables may not always be measured on the same scale, the first step in PCA is to transform the original variables $\left(X_1, X_2, \ldots, X_p\right)$ into standard scores using the relation

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j} \qquad (1)$$

where X_{ij} is the jth variable measured on the ith individual, Z_{ij} is the standard score of the jth variable measured on ith individual on the jth variable, $\bar{X}_j$ is the mean of the jth variable measured on the ith individual and S_j is the standard deviation of the jth variable measured on the ith individual.

Notably, the values of $\bar{X}_j$ and S_j are respectively calculated using the relations:

$$\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}, \; j = 1, 2, \ldots, p \qquad (2)$$

and

$$S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)}, \; j = 1, 2, \ldots, p \qquad (3)$$

Next, the correlation matrix (R) of the variables is calculated from the standard scores by calculating the sum of squares and cross products. The result gives an $n \times p$ symmetric matrix of intercorrelations among the p variables with unity along the main diagonal of the correlation matrix. Consequently,

$$R = \left(r_{ij}\right) = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}_{(p \times p)} \qquad (4)$$

An alternative means of obtaining the correlation matrix is to first obtain the covariance matrix

$$S = \left(s_{ij}\right) = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}_{(p \times p)} \qquad (5)$$

and then determine the elements of (4) from the elements of (5) using the relation

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}\, s_{jj}}} = \frac{s_{ij}}{s_i s_j} \qquad (6)$$

where $s_{ii} = \dfrac{\sum_{i}^{n}\left(X_{ri} - \bar{X}_i\right)^2}{n-1}$, $s_{jj} = \dfrac{\sum_{j}^{n}\left(X_{rj} - \bar{X}_j\right)^2}{n-1}$

and $s_{ij} = \dfrac{\sum_{i}^{n}\sum_{j}^{n}\left(X_{ri} - \bar{X}_i\right)\left(X_{rj} - \bar{X}_j\right)}{n-1}, i \neq j$

It may be noted that the correlation matrix shows how each of the variables is associated with each other. A high correlation indicates that two variables are associated and will be probably grouped together by the principal component analysis. Variables with low correlations usually will not have high loadings on the same component. Essentially, the test of hypothesis is performed to determine the significance or otherwise of the correlations.

The next task is to determine the eigenvalues $\left(\lambda_1, \lambda_2, \ldots, \lambda_p\right)$ and corresponding eigenvectors $\left(a_1, a_2, \ldots, a_p\right)$ of the correlation matrix (R). The eigenvalues of (R) are obtained by finding the roots of the characteristic equation

$$|R - \lambda I| = 0 \qquad (7)$$

where I is a $p \times p$ identity matrix and λ is an eigenvalue of the correlation matrix

Similarly, the eigenvectors corresponding to λ is obtained by solving the equation

$$\left(R - \lambda I\right)a = 0 \qquad (8)$$

where a is an eigenvector corresponding to the eigenvalue λ of (R). The eigenvector of a

Further, we use obtain the normalizing eigenvector of (R) by solving the orthogonal condition

$$a^1 a = 1 \qquad (9)$$

In view of (7), (8) and (9) respectively, the model for the PCA is formulated as follows

$$\begin{aligned}
Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
\vdots \quad & \quad \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \\
Y_p &= a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
\end{aligned} \qquad (10)$$

where $Y_1, Y_2, \ldots, Y_p$ are the first, second,…, pth principal components, written as the linear combinations of the original variables $X_1, X_2, \ldots, X_p$ and $a_{j1}, a_{j2}, \cdots a_{jp}$ are the elements of the corresponding normalized eigenvectors.

2.2 Criteria for the number of components to be extracted

Once we have obtained the principal components, there is need to decide how many components to retain. One of the celebrated devices for extracting principal components is the scree plot. The scree plot is a plot of the eigenvalues λ_j, $j = 1, 2, \ldots, p$ against all the components j, when the eigenvalues are ordered from the largest to the smallest. From the scree plot, we look for an elbow (bend). The number of components is taken to be the point at which the remaining eigenvalues are relatively small and about the same size (Johnson and Wichern, 2007).

Another rule of thumb is to retain components whose eigenvalues λ_j, $j = 1, 2, \ldots, p$ are greater than unity, or equivalently, only those components which, individually, explain at least a proportion $1/p$ of the total variance in the dataset. Also, we can retain all components that combine to account for at least 70% of the total variation in the original data.

To compute the proportion of variance in the original data explained by the jth principal component, the following formula is utilized:

$$p_j = \frac{\lambda_j}{\sum_{j=1}^{p}\lambda_j} \qquad (11)$$

2.3 Criteria for the significance of component loadings

Component loading represents the correlation between an original variable and its component. To decide which component loadings are worth considering, one is required to adopt a rule of thumb, which suggests that component loadings greater 0.30 should be considered significant; loadings of 0.40 are considered more important; and if the loadings are 0.50 or greater, they are considered very significant , provided the sample size is 50 or larger (Hair Jr., 1992).

III. Results And Discussion

3.1 Main Results

3.1.1 *Mean vector, covariance matrix and correlation matrix of the trouser measurements*

The variables used for this study are defined as follows:

X_1 : Waist measurement

X_2 : Length measurement

X_3 : Lap measurement

X_4 : Knee measurement

X_5 : Flap measurement

X_6 : Base measurement

Applying (2) to the research data, one obtains the following mean vector

$$\bar{X}^T = \begin{bmatrix} 32.672 & 40.920 & 26.530 & 19.610 & 10.940 & 18.320 \end{bmatrix}$$

Using (4) and (5) on the research data, one obtains the covariance and correlation matrix displayed as follows:

$$\boldsymbol{S} = \begin{pmatrix} 6.226139 & 4.257918 & 1.740653 & 0.638857 & 0.457469 & 0.460163 \\ 4.257918 & 22.513878 & 4.323878 & 0.585510 & 0.505306 & -0.193265 \\ 1.740653 & 4.323878 & 5.045000 & 0.491531 & 0.160000 & 0.020816 \\ 0.638857 & 0.585510 & 0.491531 & 0.574388 & 0.180204 & 0.147755 \\ 0.457469 & 0.505306 & 0.160000 & 0.180204 & 0.363673 & 0.147143 \\ 0.460163 & -0.193265 & 0.020816 & 0.147755 & 0.147143 & 0.579184 \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} 1.000 & 0.360 & 0.311 & 0.338 & 0.304 & 0.242 \\ 0.360 & 1.000 & 0.406 & 0.163 & 0.177 & -0.054 \\ 0.311 & 0.406 & 1.000 & 0.289 & 0.118 & 0.012 \\ 0.338 & 0.163 & 0.289 & 1.000 & 0.394 & 0.256 \\ 0.304 & 0.177 & 0.118 & 0.394 & 1.000 & 0.321 \\ 0.242 & -0.054 & 0.012 & 0.256 & 0.321 & 1.000 \end{pmatrix}$$

3.1.2 *Hypothesis test on the correlation between the trouser measurements*

The pairwise correlations and their corresponding p-values are presented in Table 2.

Table 2: Test of significance of the correlation between two variables

S/No.	Pair of variables (X_i, X_j)	Correlation between variable (r_{ij})	P-value	Remarks 0.05 level
1	(X_1, X_2)	0.360	0.010	*Not Significant*
2	(X_1, X_3)	0.311	0.028	*Not Significant*
3	(X_1, X_4)	0.338	0.016	*Not Significant*
4	(X_1, X_5)	0.304	0.032	*Not Significant*
5	(X_1, X_6)	0.242	0.090	*Significant*
6	(X_2, X_3)	0.406	0.003	*Not Significant*
7	(X_2, X_4)	0.163	0.259	*Significant*
8	(X_2, X_5)	0.177	0.220	*Significant*
9	(X_2, X_6)	-0.054	0.712	*Significant*
10	(X_3, X_4)	0.289	0.042	*Not Significant*
11	(X_3, X_5)	0.118	0.414	*Significant*
12	(X_3, X_6)	0.012	0.933	*Significant*
13	(X_4, X_5)	0.394	0.005	*Not Significant*
14	(X_4, X_6)	0.256	0.073	*Significant*
15	(X_5, X_6)	0.321	0.02	*Not Significant*

3.1.3 *Estimates of eigenvalues of the correlation matrix*

Table 3: Results for the Extraction of Components

Components	Eigenvalues	Percentage of Variance	Cumulative Percentage of Variance
1	2.2538	37.6	37.6
2	1.2818	21.4	58.9
3	0.7166	11.9	70.9
4	0.6858	11.4	82.3
5	0.5800	9.7	92.0
6	0.4819	8.0	100.0

Figure 1 gives the plot of the eigenvalues against the number of components.

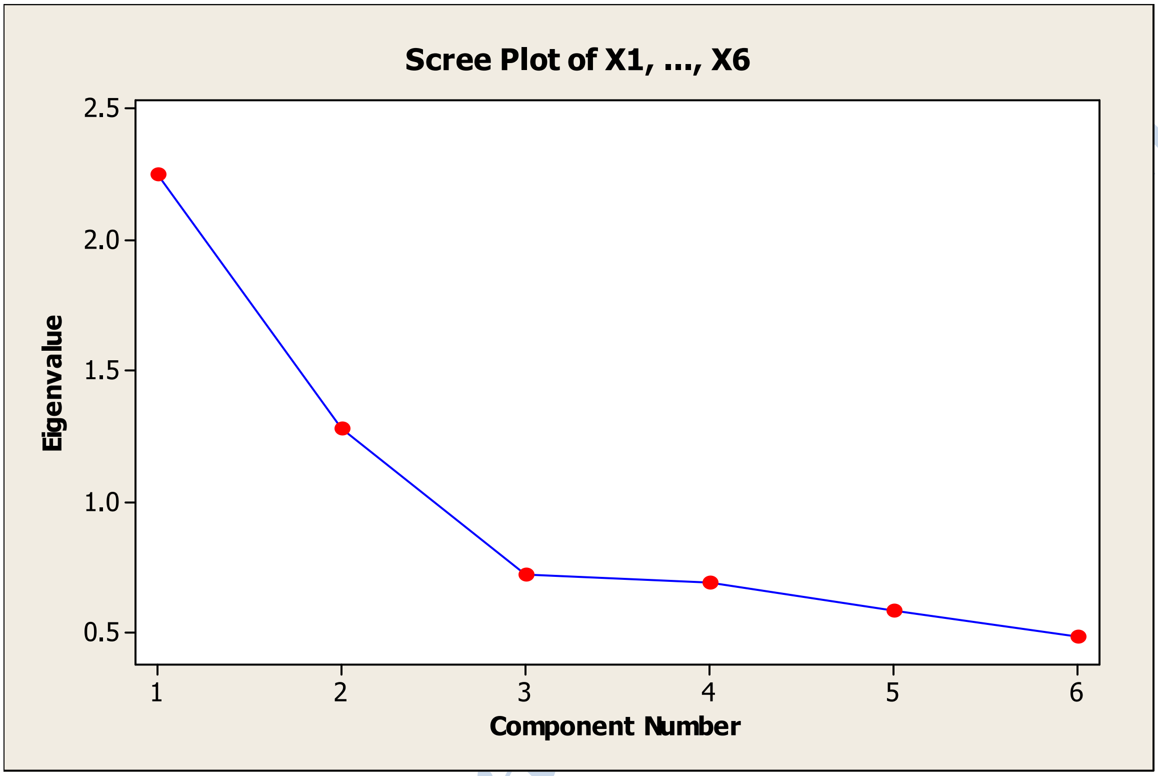

Figure 1: Scree test for Component Analysis

Table 4: Component loadings of the 6 variables

Variables	Components		
	1	2	3
X_1	0.486	0.061	-0.472
X_2	0.373	0.524	-0.329
X_3	0.390	0.458	0.294
X_4	0.461	-0.186	0.659
X_5	0.425	-0.351	0.093
X_6	0.284	-0.595	-0.372

IV. Discussion of results

Inspection of the correlation matrix and results of Table 1 reveals that 7 of the 15 correlations are significant at the 0.05 level (while 13 of the 15 correlations are significant at 0.01 level). But it is difficult to derive a complete and clear understanding of the relationships among the variables.

Table 6.4 contains the information regarding the six possible components and their relative explanatory powers as expressed by their eigenvalues. If we apply the eigenvalue criterion, then two components will be retained. The scree plot (Figure 1), however, indicates that three components may be appropriate. In viewing the third component, its low value (0.7166) relative to the eigenvalue criterion of 1.0 precluded its inclusion. Since its eigenvalue is quite high and could be rounded up to 1.0, then we include the third component as well. Still the three components retained represent 70.9 percent of the variance of the six variables under study. Thus, most of the data structure can be captured in two or three underlying dimensions. The remaining principal components account for a very small proportion of the variability and are probably unimportant.

From Table 6.4, we notice that the first principal component has variance (eigenvalue) of 2.2538 and accounts for 37.6% of the total variance. The coefficients listed under PC1 show how to calculate the principal component scores:

$$Y_1 = 0.486X_1 + 0.373X_2 + 0.390X_3 + 0.461X_4 + 0.425X_5 + 0.284X_6$$

It should be noted that the interpretation of the principal components is subjective; however, obvious patterns emerge quite often. For instance, one could think of the first principal component as representing an overall basic trouser measurement, because the coefficients of these terms have the same sign. It is obvious that all loadings are significant except the loading for variable six (base measurement).

Similarly from Table 4, it is observed that the second principal component has variance of 1.2818 and accounts for 21.4% of the data variability. The second principal component is calculated from the original data using the coefficients listed under PC2 and we obtain:

$$Y_2 = 0.061X_1 + 0.524X_2 + 0.458X_3 - 0.186X_4 - 0.351X_5 - 0.595X_6$$

This component could be thought of as contrasting waist measurement, Length measurement and Lap measurement with Knee measurement, Flap measurement and Base measurement to some extent. For component 2, only four loadings are significant.

Finally from Table 4, it is observed that the third principal component has variance of 0.7166 and accounts for 11.9% of the data variability. It is calculated from the original data using the coefficients listed under PC3:

$$Y_3 = -0.472X_1 - 0.329X_2 + 0.294X_3 + 0.659X_4 + 0.093X_5 - 0.372X_6$$

Component 3 shows that waist, length and base measurements can be grouped together while lap, knee and flap measurements should be taken together during the course of trouser production. Obviously, four loadings in component 3 are significant.

V. Conclusion

In this paper, we have demonstrated the use of the multivariate statistical technique of principal component analysis in determining the underlying structures of measurements taken by tailors to sew a pair of trousers. Interestingly, the measurements can be presented adequately in three dimensions because we obtained only three principal components. The first dimensions shows that waist, length, lap, knee, and flap are significant for sewing a pair of trousers. The second dimension shows that four measurements (length, lap, flap and base) could be taken together during trouser production and the third dimension suggests the use of four measurements (waist, length, knee, base) in sewing a pair of trousers.

References

Anyadike R.N.C. (2009). *Statistical Methods for the Social and Environmental Sciences*. Spectrum Books Limited, Nigeria.

Enogwe S.U. (2011). *Principal Component Analysis of Tailors measurements taken to sew a pair of trousers: A case study of ten selected tailors in Aba, Abia State*. Unpublished HND Project, Department of Mathematics/Statistics, Federal Polytechnic Nekede, Owerri.

Hair, Jr., J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1992). *Multivariate Data Analysis with Readings*, Third Edition, Macmillan Publishing Company, New York.

Johnson and Wichern (2007). *Applied Multivariate Statistical Analysis, Sixth Edition*, Englewood Cliffs, New Jersey.

Onyeagu, S.I. (2003). *A First Course in Multivariate Statistical Analysis*, Mega Concepts, Awka, Anambra State