Modelling relationship between students' pre and post-admission performances

Olayiwola O. M[1], Bisira, H. O[2], Wale-Orojo O. A[1], Aderele O. R[2] and Adeniyi, M. O[2]

[1]Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria.

[2]Department of Mathematics and Statistics, Lagos State Polytechnic, Ikorodu, Lagos Nigeria.

ABSTRACT

Many researchers have carried out research on students' academic performances in the University system, but there is dearth of information about models which focus on the relationship between students pre and post admission performances. This paper was therefore designed to model relationship between students' pre and post admission performances. Information on pre admission performances (Olevel, Jamb and Post-Jamb results) and post-admission performances (100L – 400L results) of students' in Statistics Department, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria was collected from students' file. Correlation matrix and Canonical Correlation analysis were used to know the degree of relationship that exists between the pre and post admission performances. The principal component analysis was employed to reduce the multidimensional data. Scree plot was used to determine the spread of the trend of the components and bi plot was used to determine the degree of closeness of the students' pre and post admission performances. There is no relationship between pre and post admission performances. Also, no strong relationship among pre admission performances, while the relationship among post admission performances is very high. Post admission performances are highly related to students' CGPA. The proportion of variance accounted for by the first, second and third principal Components are 50.7542%, 16.5712% and 15.6224% respectively with cumulative proportion of 82.95%. The first, second and third components are chosen. The seven components were reduced to three. Post admission performances are closely related and stand as the determinant of students' class of degree.

Introduction

Models of school differences in educational achievement typically assess the progress that pupils make between two test occasions and attempt to assess the extent to which variation between pupils is attributable to differences between schools. These models are referred to as value added models and the current preferred practice is estimation using multilevel models (Aitkin and Longford 1986, Raudenbush and Bryk 1986 and Goldstein et al 1993. Students mobility and neighbourhood effects are often discussed as important potential influences on educational achievement (Office for Standards in Education 2002, Department of Education and Science 2003, Association of London Government 2005, Greater London Authority 2005. Where studies look at the effect of whether a pupil has moved schools or not they find an overall negative association (Yang et al 1999), but this has not been explored for different types and timings of moves. Goldstein et al 2007 treated pupils as belonging to only their final schools and ignore the contribution of earlier schools that were attended. The studies that have looked for neighbourhood effects on education achievement have not been able additionally to model pupil movements (Garner and Raudenbush 1991, Fielding et al 2006). Research into pupil mobility has been held back by both a lack of data on pupil movements and also by the absence of appropriate methodology. The recently established national pupil data base in England and the development of cross-classified and multiple-membership multilevel models now make it possible to analyse a wide range of complex non-hierarchical data structures in models of educational achievement (Fielding and Goldstein 2006, Rasbash and Browne 2001, 2008)

Methodology

Principal Component Analysis

Suppose that x is a vector of p random variables, and that variances of the p random variables and the structure is very simple, it will often not be very helpful to simply look at the p variances and all of the $\frac{1}{2}[p(p-1)]$ correlations and covariance. An alternative approach is to look for a few ($<p$) derived variables that preserve most of the information given by these variances, correlations and covariance.

Although principal component analysis does not ignore correlations and covariance, it concentrates on variances. The first step is to look for a linear function $\alpha^{1}x$ of the elements of x having maximum variance where α_1 is a vector of p constants $\alpha_{11}, \alpha_{11},..., \alpha_{11}$ and (') denote the transpose, so that

$$\alpha_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + ... + \alpha_{1p}x_p = \sum_{j=1}^{p}\alpha_{1j}x_j$$

Tele:
E-mail addresses: laniyimathew@yahoo.com

Then we look for a linear function $\alpha_2^1 x$, uncorrelated with the $\alpha_1^1 x$, having maximum variances, and so on, so that the kth stage linear function of $\alpha_k' x$ is found that has maximum variance subject to being uncorrelated with $\alpha_1' x, \alpha_2' x, ..., \alpha_{k-1}' x$. The kth derived variable $\alpha_k' x$ is the kth principal component up to p PC could be found, but it is hoped that in general, that most of the variation of x will be accounted for by m PCs, where $m<p$. the reduction is complexity achieved by transforming the original variable to PCs.

Consider for the moment, the case where the vector of random variables x has a known covariance matrix $\sum$. This is a matrix whose $(i,j)^{th}$ element is known, covariance between i^{th} and j^{th} element of x when $i \neq j$, the variance of the jth element of x when $i = j$.

To derive the PC using first $\alpha_k' x$ ' the vector α_1 maximizes

$\text{var}[\alpha_1' \sum \alpha_1]$ subject to $\alpha_1' \alpha_1 = 1$, the standard approach is to use the technique LaGrange Multiplier.

Maximize $\alpha_1' \sum \alpha_1 - \lambda(\alpha_1' \alpha_1)$,

λ is the LaGrange multiplier.

Differentiation with respect to α_1' gives

$$\sum \alpha_1' - \lambda \alpha_1 = 0 \quad \text{or} \quad (\sum - \lambda I_p)\alpha_1 = 0,$$

I_p is the ($p \times p$) identity matrix. Thus, λ is an eigenvalue of $\sum$ and α_1 is the corresponding eigenvector. To decide which of the p eigenvectors gives $\alpha_1' x$ with maximum variance, we should know that the quantity to be maximized is

$$\alpha_1' \Sigma \alpha_1 = \alpha_1' \lambda \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda$$

So λ must be as large as possible. Thus, α_1 is the eigenvector corresponding to the largest value of $\sum$, and $\text{var}(\alpha_1' x) = \alpha_1' \Sigma \alpha_1 = \lambda$, which is the largest eigen-value. The second PC, $\alpha_2' x$, maximizes $\alpha_2' \Sigma \alpha_2$ subject to the uncorrelated matrix with $\alpha' x$, or equivalently subject to $\text{cov}(\alpha_1' x, \alpha_2' x)=0$, where $cov(x,y)$ denotes the covariance between random variable x and y.

But

$$\text{cov}(\alpha_1' x, \alpha_2' x) = \alpha_1' \Sigma \alpha_2 = \alpha_2' \Sigma \alpha_1 = \alpha_2' \Sigma \alpha_1 = \alpha_2' \lambda_1 \alpha_1' = \lambda_1 \alpha_2' \alpha_1$$

Thus, any of the equations

$\alpha_1' \Sigma \alpha_2 = 0, \quad \alpha_2' \Sigma \alpha_1 = 0,$

$\alpha_1' \alpha_2 = 0$, $\alpha_2' \alpha_1 = 0$

Could be used to specify zero correlation between $\alpha_1' x$ and $\alpha_2' x$. Choosing the last of these equations, and noting that a normalization constraint is necessary, the quantity to be maximized is

$$\alpha_2' \Sigma \alpha_2 - \lambda(\alpha_2' \alpha_2 - 1) - \phi \alpha_2' \alpha_1,$$

where λ and ϕ are LaGrange Multipliers.

Differentiating with respect to α_2 gives $\Sigma \alpha_2 - \lambda \alpha_2 - \phi \alpha_1 = 0$ and multiplying this equation on the left by α_2' gives

$\alpha_2' \Sigma \alpha_2 - \lambda \alpha_2' \alpha_2 - \phi \alpha_1 = 0$ Which since the first two terms are zero and $\alpha_1' \alpha_1 = 1$ reduces to $\phi = 0$. Therefore $(\sum - \lambda I_p)\alpha_1 = 0$ is once more eigenvalue of $\sum$ and α_2 the corresponding eigenvector.

Biplots

Biplots similarly provide plots of the n observations, but simultaneously they give plots of the relative positions of the p variables in two dimensions. Furthermore, superimposing the two types of plots provides additional information about relationships between variables and observations not available in either individual plot.

Eigen Value And Eigen Vector

For a square matrix $\mathbf{A}$ of order n the scalar λ is said to be the eigenvalue (or characteristic root or simply a root) of $\mathbf{A}$ if $\mathbf{A} - \lambda \mathbf{I}_n$ is singular. Hence the determinant of $\mathbf{A} - \lambda \mathbf{I}_n$ must equal zero

$|\mathbf{A} - \lambda \mathbf{I}_n| = 0$ (1)

Equation (1) is called the characteristic or eigenequation of the square matrix $\mathbf{A}$ and is An n-degree polynomial in λ with eigenvalues (characteristic roots) $\lambda_1, \lambda_2, \ldots, \lambda_n$. If some subset of the roots are equal, say $\lambda_1 = \lambda_2 = \cdots = \lambda_m$, where $m < n$, then the root is said to have multiplicity m.

Hence, there exist nonzero vectors $\mathbf{p}i$ such that

$(\mathbf{A} - \lambda_k \mathbf{I}_n) \mathbf{p}_i = \mathbf{0}$ for $i = 1, 2, \ldots, n$ (2)

The vectors $\mathbf{p}_i$ which satisfy (2) are called the eigenvectors or characteristic vectors of the eigenvalues or roots λ_i.

Canonical Correlation Analysis (CCA)

Canonical correlation analysis studies the relationship between a set of predictor(Independent variables) and a set of criterion (dependent) variables or between two pairs of vectors.If we have *p* predictor variables (*X*'s) and *q* criterion variables (*Y*'s) and if $p \geq q$, then the partitioned covariance matrix is shown in the following figure.

$$C = \begin{bmatrix} S_{x1}^2 & S_{x1x2} & S_{x1x3} & \cdots & S_{x1xp} & S_{x1y1} & S_{x1y2} & S_{x1y3} & \cdots & S_{x1yq} \\ S_{x2x1} & S_{x2}^2 & S_{x2x3} & \cdots & S_{x2xp} & S_{x2y1} & S_{x2y2} & S_{x2y3} & \cdots & S_{x2yq} \\ S_{x3x1} & S_{x3x2} & S_{x3}^2 & \cdots & S_{x3xp} & S_{x3y1} & S_{x3y2} & S_{x3y3} & \cdots & S_{x3yq} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{xpx1} & S_{xpx2} & S_{xpx3} & \cdots & S_{xp}^2 & S_{xpy1} & S_{xpy2} & S_{xpy3} & \cdots & S_{xpyq} \\ \hline S_{y1x1} & S_{y1x2} & S_{y1x3} & \cdots & S_{y1xp} & S_{y1}^2 & S_{y1y2} & S_{y1y3} & \cdots & S_{y1yq} \\ S_{y2x1} & S_{y2x2} & S_{y2x3} & \cdots & S_{y2xp} & S_{y2y1} & S_{y2}^2 & S_{y2y3} & \cdots & S_{y2yq} \\ S_{y3x1} & S_{y3x2} & S_{y3x3} & \cdots & S_{y3xp} & S_{y3y1} & S_{y3y2} & S_{y3}^2 & \cdots & S_{y3yq} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{yqx1} & S_{yqx2} & S_{yqx3} & \cdots & S_{yqxp} & S_{yqy1} & S_{yqy2} & S_{yqy3} & \cdots & S_{yq}^2 \end{bmatrix} = \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix}$$

The problem is to find a vector a and vector *b* such that the correlation between $a^t X \text{ and } b^t Y$ (*canonical variates*--new variables that are obtained by using linear combinations of the original variables) is maximal.

$$r(a^T X, b^T Y) = \frac{a^T C_{XY} b}{[(a^T C_{XX} a)^{0.5} (b^T C_{YY} b)^{0.5}]}$$

Where $C_{XY} = C_{XY}^T$ between set covariance matrix

$C_{XX} =$ covariance matrix of variables $(p \times p)$

$C_{YY}=$ covariance matrix of variables $(q \times q)$ in *Y*

The problem is to maximize $r(a^T X, b^T Y)$ such that $a^T C_{XX} a = 1 =$ and $b^T C_{YY} = 1$ (constraints).

This reduces to a problem of solving the following canonical equations:

$C_{XX}^{-1} C_{XY} C_{YY}^{-1} C_{YX} - \lambda I_P) a = 0$, where *0* and *a* are column vectors $(p \times 1)$; I $_{p \text{ is}}$ the identity matrix of size *p* ×*p*; λ is an eigenvalue (a single value, *1×1*), starting with the largest one.; and $C_{YY}^{-1} C_{YX} C_{XX}^{-1} C_{XY} - \lambda I_P) b = 0$ where *0* and *b* are column vectors $(q \times 1)$; I $_q$ is the identity matrix identity matrix of size $(q \times q)$; λ;

And

$|C_{XX}^{-1} C_{XY} C_{YY}^{-1} C_{YX} - \lambda I_P) a| = 0$
$|C_{YY}^{-1} C_{YX} C_{XX}^{-1} C_{XY} - \lambda I_P) b| = 0$

If we standardize the variables (note: cannot do this is some of the variables aredummy variables representing nominal or ordinal variables), then the equationscould be expressed in terms of the correlation matrices:

$R_{XX}^{-1} R_{XY} R_{YY}^{-1} R_{YX} - \lambda I_P) a = 0$
$R_{YY}^{-1} R_{YX} R_{XX}^{-1} R_{XY} - \lambda I_P) b = 0$

And

$$|R_{XX}^{-1} R_{XY} R_{YY}^{-1} R_{YX} - \lambda I_P) a| = 0$$

$|R_{YY}^{-1} R_{YX} R_{XX}^{-1} R_{XY} - \lambda I_P) b| = 0$

The eigenvalues of these equations are the *squared canonical correlation coefficients*. This is similar to the coefficient of determination (R^2 value) for multiple linear regression analysis. The eigenvectors associated with the eigenvalues are the vectors of coefficients *a* and *b* called *canonical weights*.

$$a = \frac{R_{XX}^{-1} R_{XY} b}{\sqrt{\lambda}} \qquad b = \frac{R_{YY}^{-1} R_{YX} a}{\sqrt{\lambda}}$$

At most *p* (since $q \geq p$) canonical variates can be extracted. Successive canonical variates extracted have canonical correlations decreasing in magnitude and are uncorrelated with each other.

Discussion of results

There is no relationship between pre admission performance (Olevel, JAMB and Post JAMB) and post admission performance (100L, 200L, 300L and 400L performances). Similarly for relationship among pre admission performances, while the relationship among post admission performances is positive and strong. Post admission performances are strongly related to the CGPA of the students, while pre admission performances are not (From table 1)

The proportion of variance accounted for by the first, second and third principal Components are 50.7542%, 16.5712% and 15.6224% respectively with cumulative proportion of 82.95%. This implies that the first, second and third principal components are sufficient to explain the students' CGPA (From table 2)
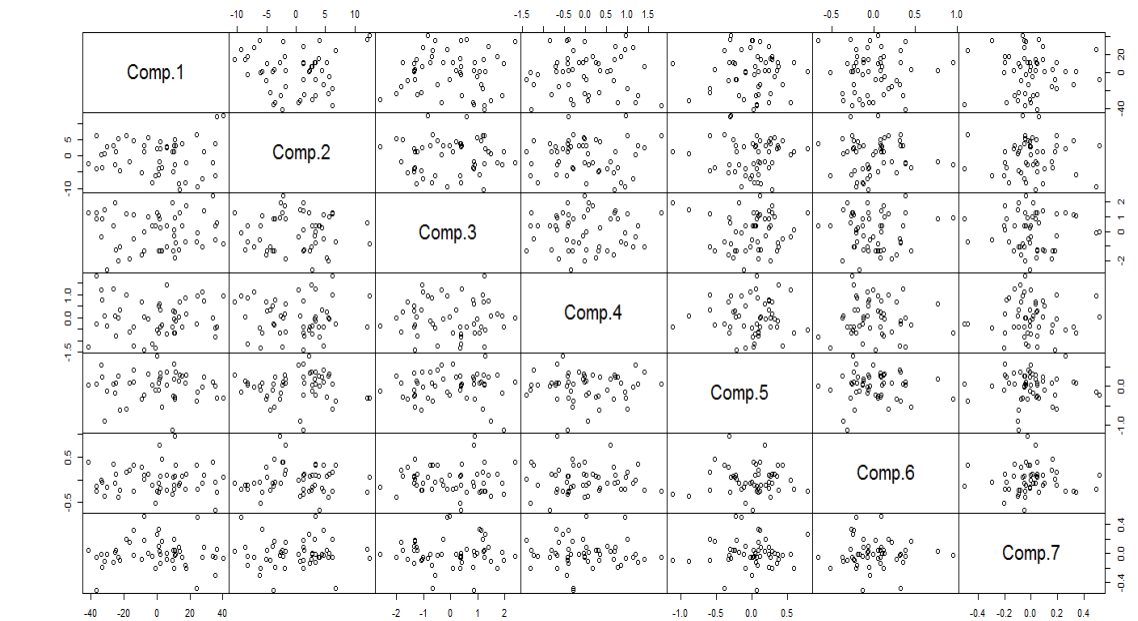

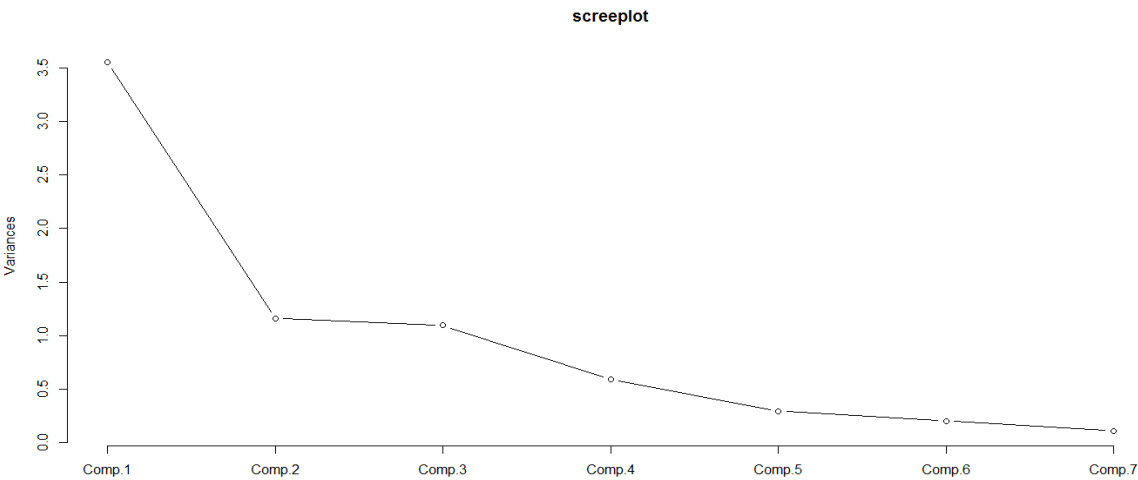Figure 1: Pair plot of the principal components
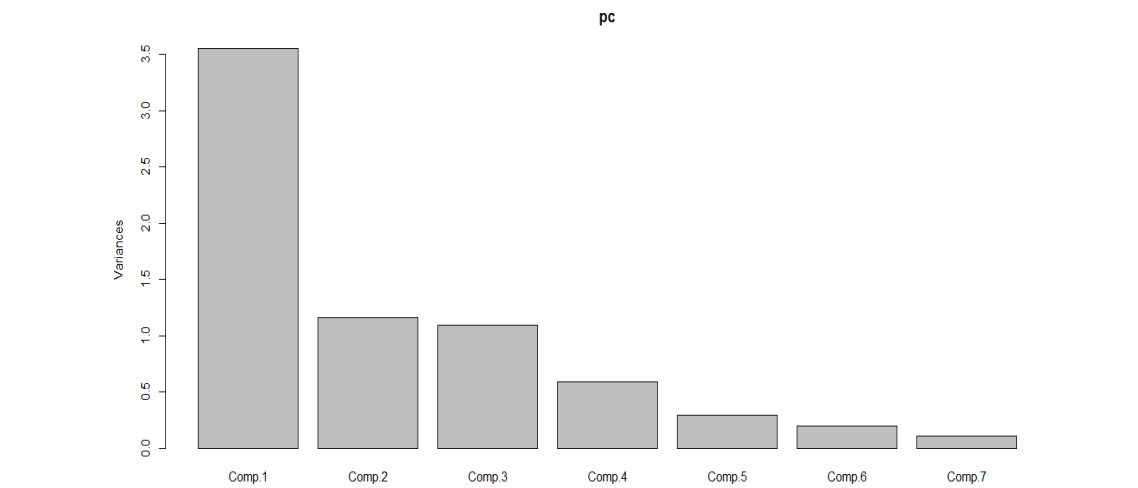

Figure 2: Scree Plot Of The Principal Components

Figure 3. Principal Components Bar Chat

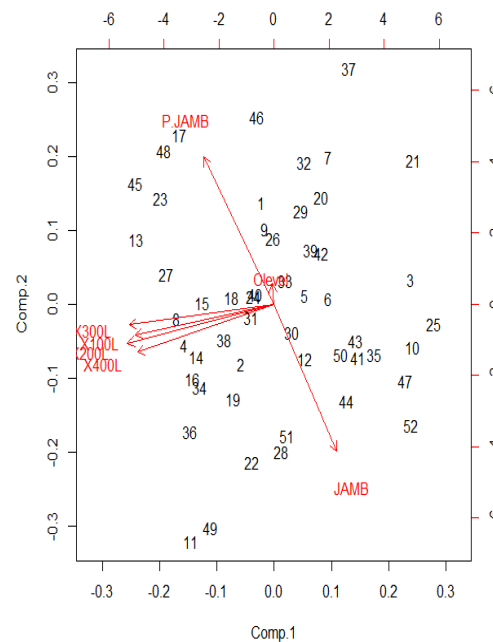**Figure 4: Principal Components Bi-Plot**

Table 1: Correlation Matrix

	Olevel	JAMB	P.JAMB	100L	200L	300L	400L	CGPA
Olevel	1.0000	0.1107211	0.1501844	0.0699603	-0.0554020	-0.07780	0.08772771	-0.008198
JAMB		1.0000	-0.38667	-0.208181	-0.1866863	-0.3083229	-0.2286422	-0.2707760
P.JAMB			1.0000	0.3014470	0.3192679	0.30833121	0.16732963	0.303616260
100L				1.0000	0.8196233	0.70841452	0.71451165	0.861060219
200L					1.0000	0.84865850	0.75651953	0.933186608
300L						1.0000	0.77480792	0.943977668
400L							1.0000	0.889219163
CGPA								1.0000

Table 2: Principal Component Analysis

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	1.8848861	.077025 1	.0457384	0.76694744	0.54250377	0.44883876	0.33117762
Proportion of Variance	0.5075420	.165712	0.1562241	0.08402977	0.04204433	0.02877946	0.01566837
Cumulative Proportion	.507542	0.673254	0.8294781	0.91350783	0.95555216	0.98433163	1.00000000

Table 3: Significant Loading Of The Principal Components

	Component 1	Component 2	Component 3
OLEVEL		0.918	-0.315
JAMB	0.209	-0.664	0.304
POST JAMB	-0.237	0.666	0.196
100 L	-0.465	-0.139	
200 L	-0.490		
300 L	-0.484	-0.177	-0.106
400 L	-0.457	-0.219	

Figure 1 shows the pair plot for the principal component analysis of the seven variables. It shows the degree of spread of the variables. The spread has its range from the negative to the positive values. From the Scree plot in figure 2, it shows that spread of the trend of the components. The best components are often greater or equals to 1. Hence, first, second and third components are chosen. Therefore, it reduced the seven components to three components. From figure 3, It is clearly shown that the first, second and third components are equal or greater than 1. So, they are the best components for the principal component analysis but the first component is the best component. The bi plot in figure 4 shows the degree of closeness of the students' performance. It is observed that the 100L, 200L, 300L and 400L results are closely related and have strong degree of relationship.

From table 3, it alsowas observed that the best components to be chosen are the first, second and third components and this leads to the formulation of the PCA model below:

$$Comp1 = 0.209JAMB - 0.237P.JAMB - 0.465X100L - 0.490X200L - 0.484X300L - 0.457X400L$$

$$Comp2 = 0.918Olevel - 0.664JAMB + 0.666P.JAMB - 0.139X100L - 0.177X300L - 0.219X400L$$

$$Comp3 = -0.315Olevel + 0.304JAMB + 0.196P.JAMB - 0.106X300L$$

Conclusion

There is no relationship between pre and post admission performances. Also, no strong relationship among pre admission performances, while the relationship among post admission performances is very high. Post admission performances are highly related to students' CGPA. The proportion of variance accounted for by the first, second and third principal Components are 50.7542%, 16.5712% and 15.6224% respectively with cumulative proportion of 82.95%. The first, second and third components are chosen. Therefore, it reduced the seven components to three components. Post admission performances are closely related and stand as the determinant of students' class of degree.

References

❖ Aitkin, M. and Longford, N. (1986) Statistical modeling issues in school effectiveness studies (with discussion). J.R. Statist. Soc. A., 149, 1-42

❖ Sssociation of London Government (2005) Breaking point: examining the distruption caused by pupil mobility. Association of London Government, London. (Available from http://www.londoncouncils.gov.uk/london%20councils/local%20government%20finance /local%20government%20finance%20publications/ALGmobilityreportfina.pdf)

❖ Department of Education and Science (1991) The parents Chartered, London:Department of Education and Science.

❖ Fielding, A., Thomas, H., Steele, F., Browne., W., Leyland, A., Spencer, N. and Davison, I. (2006). Using cross-classified multilevel models to improve estimate of the determination of pupil attainment: a scoping study. Research Report, School of Education, University of Birmingham, Birmingham. (Available from http://www.socscistaff.bham.ac.uk/fielding/Dfes_scoping_report_school_of_Ed_version_August06.pdf)

❖ Garner, C.L., and Raudenbush, S.W. (1991) Neighborhood effects on educational attainment: a multilevel analysis. Social Educ., 64, 251 – 262

❖ George Leckie (2009) The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. J.R. Statist. Soc. A, 172, 537 – 554.

❖ Goldstein, H., Burgess, S. and McConnell. B (2007) Modelling the effect of pupil mobility on school differences in educational achievement. J. R. Statis. Soc A, 170, 941- 954

❖ Goldstein, H. Rasbash. J,. Yang. M., Woodhouse, G. Pan, H., Nuttall, D, and Thomas, S. (1993). A multilevel analysis of school examination results. Oxf. Rev. Edu., 19, 425 – 433

❖ Greater London Authority (2005) Moving home and changing school – 1: widening the analysis of pupil mobility. Briefing 2005/32. Data Management and Analysis Group, Greater London Authority, London. (Available from www.bristol.ac.uk/Depts/CMPD/PLUG/publications/ewens.pdf)

❖ Office for Standards in Education (2002)Management pupil mobility. Office for Standards in Education, London. (Available from http://preview.ofsted.gov.uk/assets/65.pdf)

❖ Rasbash. J and Browne, W. (2001). Modelling non-hierarchical structures. In m ultilevel Modelling of Health Statistics (eds A.H. Leyland and H. Goldstein). Chichester, Wiley

❖ Rasbash., J,. and Browne, W. J. (2008). Non-hierarchical multilevel models. In handbook of multilevel analysis (eds, J de Leeuw and E. Meijer). New York: Springer

❖ Raudenbush, S. and Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociol Educ., 59, 1- 17.

❖ Yang, M., Goldstein, H., Rath, T. and Hill, N. (1999). The use of assessment data for school improvement purposes. Oxf. Rev. Edu., 25, 459 - 483