Journal of Management
Vol. III, No. 1, October 2005, pp 58-62

Growth Performance of Shorea Seedlings at Kalutara District in Sri Lanka

A. Jahufer*

Abstract

Principal Component Analysis (PCA) is one of the oldest and most widely used multivariate statistical techniques. The original purpose of PCA was to reduce a large number (p) of variables to a much small number (m) of principal components whilst retaining as much as possible of the variation in the p-original variables. The technique is especially useful if m<<<p, and if the m principal components can be readily interpreted.

The main objective of this research study is to find the growth performance of seedlings of seven rain forest canopy, dominant Shorea species, in Kalutara district within the humid zone of Sri Lanka for a period of 24 months.

A research was carried out by Prof. I.A.U.N. Gunetilleka and Prof. C.V.S. Gunetilleka (1989) to collect the data from 336 Shorea seedlings which were transplanted at Kalutara district in 125m altitude elevations. At the end of the experiment period only 112 seedlings were chosen at random for the analysis. The measurements of constituent parts such as stem weight, leaf weight, tap root weight, fine root weight, leaf number and total dry mass of those randomly selected seedlings were recorded in the process. The above two researchers analyzed the recorded data using the ANOVA and GLM procedure of SAS in one-way ANOVA models for each of the attributes of these constituent parts, to be measured separately to compare the performance of species grown at this district.

In this study, all the constituent parts of one species is combined using the principal component analysis, and used this new combined new variable to compare the performance of seven species grown at this district. To analyze this combined and reduced response variable (data) ANOVA procedure of SAS was used. Tukey's studentized range method was used for multiple comparison of means among species at the 5% level of significance.

Keywords : Principal Component Analysis; Eigenvalues; Eigenvectors; ANOVA, GLM, Tukey's Studentized Mean Comparison.

Introduction

The species that were selected for this research are: Affinis, Trapezifolia, Gardneri, Cordifolia, Disticha, Megistophylla and Worthingtonii. The first two are soft wooded, peeler species, used for plywood manufacture and other five are, medium hardwoods, important in the construction industry.

Principal Component Analysis (PCA)

The PCA technique was originally introduced by Pearson(1901) and independently by Hotlling(1933). PCA is concerned with explaining the variance-covariance structure through a few linear combination of the original variables, that can be used to summarize a data set, losing in the process as little information as possible. The derived variables may be used in a variety of ways.

Let the random vector $X'=[X_1, X_2, \ldots X_p]$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$. Then the following linear combination can be defined.

$$Y_i = e_i'X = e_{1i}X_1 + e_{2i}X_2 \ldots + e_{pi}X_p \quad i = 1,2,3, \ldots p$$

Where X_i, i = 1,2,3, . . . p, are the original variables, Y_i, i = 1,2,3, . . . p are the principal

*A. Jahufer, is a Senior Lecturer in the Department of Mathematical Sciences, South Eastern University of Sri Lanka, Sri Lanka.

components and e_i' is the i^{th} eigenvector of Σ.

The total population variance is =

$$\sum_{i=1}^{p} \text{var}(Xi) = \sum_{i=1}^{p} \text{var}(Yi)$$

The proportion of the total variance due to i^{th} principal component is $= \lambda_i / \sum_{i=1}^{p} \text{var}(Yi)$

If only one principal component is sufficient to explain the variability of the original variables $X_1, X_2, \ldots X_p$ then consider only the first principal component. Otherwise if several (m £ p) principal components are necessary to explain the variability, then the principal component is the linear combination of the original m principal components, and it is given by the equation as:

$$Z = w_1Y_1 + w_2Y_2 + \ldots + W_mY_m$$

where $Y_1, Y_2, \ldots Y_m$ are the first m principal component and $w_1, w_2, \ldots w_m$ are the weights corresponding to m principal components. The weight for w_i is given by $w_i = \lambda_i / \sum_{s=1}^{p} \lambda s$

Selecting the Number of Principal Components

The principal component analysis is seen to be a technique for transforming a set of observed variables into a *new set of variables* which are *uncorrelated* with one another. The total variation in the original p-variables is only accounted for by all p-principal components. The usefulness of these components, however, stems from their property of accounting for the variance in decreasing proportions. So the important question arises as to how many principal components are needed to provide an adequate summary of a given data set.

In this research the first two principal components were considered to get more information of the original data.

Shorea Species

Distribution of Shorea Species in Sri Lanka

In Sri Lanka, the family Dipterocarpaceae co-dominates both low(0-1000m) and mid (1000-1700m) elevation rain forest canopy. The distribution of Shorea species in Sri Lanka follows the general pattern of the family in a seasonal far east undergoing allopathic speciation in relation to altitude, local topographic variation, and possibly soil associated niche differentiation. Within Sri Lanka rain forest, several distinctive mature phase forest communities have been recognized along an altitudinal gradient from moist costal lowlands to sub-mountains regions in which a clad of eight partially sympatric Shorea species belonging to the endemic section Doona exists.

Present day natural distribution of these Shorea species may not often overlap with their potential ecological amplitude, which could only be examined by transplanting them in these different altitudes and monitoring their growth performance. This would, in addition, provide information on their growth potential useful in nature forest management and plantation silvicuture.

Test Shorea Species

The species that were selected for this research, were Specie Affinis, Specie Trapezifolia, Specie Gardneri, Specie Cordifolia, Specie Disticha, Specie Megistophylla and Specie Worthingtonii. The first two are soft wooded, peeler species, used for plywood manufacture and the other five are, medium hardwoods important in the construction industry. These species, belong to two groups locally known as the 'Beraliyas' (S. Cordifolia, S. Disticha, S. Megistophylla, S. Worthingtonii) and the 'Thiniyas' (S. Affinis, S. Trapezifolia, S. Gardneri). The Beraliyas bear larger and edible fruits that are heavily predated upon by small mammals and widely collected by villagers as a source of carbohydrate or firewood and also have a medicinal significant. The Thiniyas bear on

the other hand smaller and non-edible fruits. Both groups of species produce inflammable gum collected by the local people.

Methodology

Experimental Design and Layout

The experiment was carried out in shade houses for controlling the variable factors, constructed in an open area, free of shade casting by the surrounding vegetation on them. The experiment comprised seven main treatments (seven different Shorea species), four replicates per treatment and finally 12 seedlings per specie per replicate. In total 336 seedlings were monitored over a period of 24 months. Inside the shade house potted seedlings were arranged, in a completely randomized block design.

At the end of the experiment only 112 seedlings were chosen randomly for the analysis. The selected species constituent parts such as stem weight, leaf weight, tap root weight, fine root weight, leaf number and total dry mass were recorded at the end of the experiment.

Objective of Selecting Principal Component Analysis for This Research

The objective of this research is to find the growth performance of seven varieties of Shorea species, using their constituent parts such as stem weight, leaf weight, tap root weight, fine root weight, and leaf number. So, if and when the five constituent parts (five variables) can be reduced to a single variable, that variable represents the corresponding single species growth performance. Then it is easy to analyze and interpret the response of each species at this district under the controlled experiment. Variables of this type can be reduced to a single variable using principal component statistical technique.

Selecting Number of Principal Components

If principal component statistical technique is used for five constituent parts of each species then it can be obtained five principal components. If 70% to 90% of the total population variance can be attributed to the first one, two, or three principal components, then these components are sufficient to interpret the data to obtain the maximal information.

Therefore, initially, the first principal component was considered. The proportion of the first principal component of species Affinis, Trapezifolia, Gardneri, Cordifolia, Disticha, Megistophylla and Worthingtonii are 98.9%, 96.6%, 88.6%, 98.1%, 84%, 93.3% and 97.9% respectively, these proportion values are between 70% to 90% and above.

But in this research, to get the maximum information of the original data (maximum variability) the first two principal components were considered. The total proportion of first two principal components of the above species attributed 99.9%, 99.3%, 99.3%, 99.8%, 95.6%, 99.9% and 99.9% total population variance respectively. This means that more than 95% of original data information (variability) can be obtained when the first two principal components are taken for the analysis.

Choosing the best variable to fit model

The residuals of the reduced response variable was tested the validity of Gauss Markov's conditions. They satisfied the Gauss Markov's conditions (the residuals normally distributed with zero mean and constant variance, also the residuals are uncorrelated). Therefore the reduced response variable is the best variable to fit the model.

Results and Discussion

The one-way ANOVA model obtained using the reduced response variable and it is given in below table:

From the above ANOVA table it can be observed that the species are significantly different growth performance at this district. To find the significant growth performance of species Tukey's Studentized mean separation was used at 5% significance level. The output of Tukey's mean separation is given in the Table below:

Table 3.1: ANOVA table of reduced response variable

General Linear Models Procedure
Class Level Information

Class	Levels	Values
SPECIES	7	1 2 3 4 5 6 7

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	9885.9223	1647.6537	17.08	0.0001
Error	105	10131.1805	96.4874		
Corrected Total	111	20017.1028			

R-Square	C.V	Root MSE	Y Mean
0.4939	45.2742	9.8228	21.6962

From the above output table the *different letters* (Tukey Grouping) clearly indicate that the growth performance of Shorea seedlings are significantly different at this district.

The growth performance seems to be the highest in the case of species Megistophylla, and the second lower growth performance is found to be species Affinis. Likewise growth performance of the species are given in the decreasing order are: Megistophylla, Affinis, Worthingtonii, Trapezifolia, Cordifolia, Gardneri, and Disticha at this district.

Table 3.2: Tukey's mean separation for growth performance of seven Shorea seedlings

General Linear Models ProcedureTukey's Studentized Range (HSD) Test for variable: Y
Alpha= 0.05 df= 105 MSE= 96.48743
Critical Value of Studentized Range= 4.252Minimum Significant Difference= 10.44

Means with the same letter are not significantly different.

Tukey Grouping	Mean	SPECIES
A	34.785	Megistophylla
A		
B A	30.151	Affinis
B A		
B A	26.509	Worthingtonii
B		
B	22.919	Trapezifolia
B		
B C	19.926	Cordifolia
C		
D C	11.460	Gardneri
D		
D	6.124	Disticha

Conclusion

In this study all the constituent parts of one species is treated as a single variable, using the Principal Component Analysis. The random errors (residuals) of reduced response variable was used to diagnose the Gauss Markov's conditions. It was obtained that the Gauss Markov's conditions are satisfied. Therefore, the reduced response variable was used to fit into the one-way ANOVA procedure of SAS and Tukey's Studentized multiple mean comparison test was also used to find the exact difference of the species at this district.

This finding has the advantages of giving information as to the seven varieties of species instead of component parts of species, which will be helpful to have a full understanding about the seven varieties of Shorea species at Kalutara district in Sri Lanaka.

References

Anputhas, M.; Samita, S.; and Abeysiriwardena, S.De.Z. (2004). Regression Based on Multivariate Index As a Tool for Varietals Recommendation. Proceedings of the International Sri Lankan Statistical Conference: Visions of Futuristic Methodologies, 61-69.

Gunatilleke, C.V.S. and Gunatilleke, I.A.U.N. (1993). Response of Shorea (Family Dipterocarpaceae, section Doona) seedlings transplanted in different altitude within the humid zone of Sri Lanka.

Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis. 2nd Edition, John Wiley & Sons. **11**, 451-477.

Chatfied, C., and Collins, A.J., (1983). Introduction to Multivariate Analysis. Capman and Hall New York. **4**, 57-89.

John, C., Brocklebank, Ph.D, Marty Johnson, Jenny Kendall, Dominique Latour. (1995). Basic Statistics Using SAS/STAT software. SAS Institute Inc. USA.

Jolliffe, I.T., (1986). Principal Component Analysis. Springer-verlag, New York, 1-107.

Winer Donald, B.J., Brown Kenneth, R., and Michels, M., (1991). Statistical Principals in

Experimental Design. 3rd Edition, McGraw-Hill, Inc. USA. **10**, 739-837.

Douglas, C. Montgomery, and Elizabeth A. Peck, (1992). Introduction to Linear Regression Analysis. 2nd Edition, John Wiley & Sons.

Pillai, R.S.N., and Bagavathi, V., (1984). Statistics, S.Chand & Company Lid, Ram Nagar, New Delhi-110055. **11**, 316-359.

Cochran, W.G and Cox, G.M., (1992). Experimental Designs, 2nd Edition, John Wiley & Sons. 1-145.

Draper, N.R., and Smith, H., (1998). Applied Regression Analysis, 3rd Edition, John Wiley & Sons.

Moore Devid, S. (1994). The Basic Practice of Statistics, W.H. Freeman & Company, USA.

Thattil, R.O., Samita, S, Gunaratne, L.H.P (1998). Design and Analysis of Experiments, PGIA, University of Peradeniya.

Thattil, R.O., Samita, S, Gunaratne, L.H.P, Dematawewa, C.M.B, Sivayoganathan, C. Peiris, T.S.G, Jayalal Chandrasiri, C.W, (1999). Survey Design and Analysis, PGIA, University of Peradeniya.

Trenkler, G (1981). Biased Estimators in the linear Regression Model, Oelgeschlager, Gunn & Hain Publisher Inc. Cambridge, Massachusetts.