Journal Crop and Weed, 11(1):54-57(2015)

Flower production index using principal component analysis

MANOJ KUMAR, A. MAJUMDER, G. R. MANJUNATHA AND [1]K. SANJEEV

Department of Agricultural Statistics
Bidhan Chandra Krishi Viswavidyalaya, Mohanpur - 741 252, West Bengal.
[1]Department of Statistics, Mathematics and Computer application, RAU Pusa, Bihar

Received: 20-12-2014, Revised: 10-04-2015, Accepted: 15-04-2015

ABSTRACT

An indicator is a quantitative or a qualitative measure of serially observed facts that can locate relative positions in a particular area. Indicators are useful for determining trends and drawing conclusions for particular issues in policy analysis. They can also be helpful in making policy and in monitoring performance. When several indicators are compiled into a single index using a specific technique, then a composite indicator is formed. The composite indicator measures multi-dimensional concepts, which cannot be explained by a single indicator. Here, Flower Production Index (FPI) has been constructed using Principal Component Analysis (PCA) for 18 districts of West Bengal, India. In present study, the indicators like production of Rose (PR), Production of Gladiolus (PG), Production of Marigold (PM) and production of seasonal flower (PSF) have been taken. This methodology for constructing composite index can also be used in multidimensional scaling.

Keywords : Composite Indicator, flower production index, multidimensional scaling, principal component analysis

Composite Index has been constructed by several authors using different techniques. Among them, Analytic Hierarchy Process (AHP) is a technique that is being used extensively in many areas to analyze and support decisions in which many objectives (some time even competing objectives) are involved and many alternatives are available. The AHP is introduced in late 1970s. This technique is based on expert judgment and the experts of relevant field give their opinion/priorities to several alternatives to analyze and support decisions. As it is based on expert's judgment, it is subjective in nature. Using public opinion as a weighting technique, Parker (1991) developed environmental problem index. Ercot and Moran (1991) used AHP to rank a small number of municipal landfill potential sites for City of Edmonton, Alberta, Canada. Siddiqui *et al*. (1996) have used AHP in landfill siting using GIS. Narain *et al.* (1991) proposed a technique of constructing composite index and constructed a composite index of development for estimating the potential targets for the underdeveloped States to bring equity in development. This technique involves the problem of multicolinearity. Mahlberg and Obersteiner (2001) constructed human development index by using benefit of doubt as a weighting method. Ahmad *et al.* (2003) identified potential agro forestry areas by using Objective Analytic Hierarchy Process (OAHP). Narain *et al.* (2005) have estimated socio-economic development of different districts in Kerala.

The techniques used above to construct composite index is subjective in nature or/ and involves problem of multicolinearity. If multicollinearity present in the data set, the weight of one variable is added up to the weight of correlated variables. This yields poorly constructed composite index. To overcome the problem of multicolinearity, Kumar *et al.* (2013) proposed and used Principal Component Analysis (PCA) in construction of composite index and developed Agriculture Development Index of Bihar State, India. In this paper, Flower Production Index (FPI) has been constructed using PCA with available indicators for 18 districts of West Bengal State, India.

MATERIALS AND METHODS

The data on flower production has been taken from the secondary source (Directorate of Food Processing Industries and Horticulture, Govt. of W.B.) for all the 18 districts of West Bengal for the year 2007-08. District wise data on production of Rose (PR) in crore spike, production of Gladiolus (PG) in crore spike, production of Marigold (PM) in thousand mt and production of seasonal flower (PSF) also in thousand mt has been analyzed. The imputations of missing values in the data were done using the minimum value of corresponding data set.

Principal Component Analysis is generally used for data reduction technique (Dutta *et al*., 2014) as well as for the solution of multicolinearity. The principal components can be utilized in construction of composite index. All the principal components obtained from a data set can be used to construct composite index because the aim is not to reduce the data set but to rank the districts, State or countries. As principal components are uncorrelated, therefore, all principal components can be used in construction of composite index. The methodology used in construction is as under.

Email: manoj_iasri@yahoo.com

Maximum Likelihood Estimate (M.L.E.) of variance-covariance matrix (Σ) of the given data set was estimated by

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}})' \quad \ldots (1)$$

where $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_q \end{bmatrix}$

Where, q is the number of indictors / variables.

$\underline{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n}\underline{X}_i$ and n is total number of districts.

Then Correlation Matrix (CM) was obtained using above variance-covariance matrix as

$$CM = \left(\sqrt{V}\right)^{-1}\hat{\Sigma}\left(\sqrt{V}\right)^{-1} \quad \ldots (2)$$

where

V = Diagonal matrix obtained from variance-covariance matrix and $\hat{\Sigma}$ = M. L.E. of variance-covariance matrix.

Next step was to obtain principal components using eigen vectors of the estimated correlation matrix and standardized values of variables. The principal components were obtained by using the formula given below.

$$P_1 = a_{11}Z_1 + a_{12}Z_2 + \ldots + a_{1q}Z_q$$
$$P_2 = a_{21}Z_1 + a_{22}Z_2 + \ldots + a_{2q}Z_q$$
.
.
.
$$P_q = a_{q1}Z_1 + a_{q2}Z_2 + \ldots + a_{qq}Z_q$$

where

P_qs : q^{th} principal components

Table 1: Correlation matrix

	PR	PG	PM	PSF
PR	1.000	0.146*	0.767**	0.844**
PG		1.000	0.087*	0.131*
PM			1.000	0.926**
PSF				1.000

Note: *The values indicated by * is not significant at p value 0.05 and the values indicated by **is significant at p value 0.05.*

Table 2: Detection of multicolinearity

Model	Dependent variable	Independent variables	p value	R^2	VIF
1	PR	PG, PM and PSF	<0.001	0.715	3.510
2	PG	PR, PM and PSF	0.930	0.029	1.030
3	PM	PR, PG and PSF	<0.001	0.859	7.110
4	PSF	PR, PG and PM	<0.001	0.901	10.100

Table 3: Eigen values

Eigen values of the Covariance Matrix			
	Eigen value	**Proportion of explained variation**	**Cumulative**
1	2.719	0.680	0.680
2	0.977	0.244	0.924
3	0.243	0.061	0.985
4	0.061	0.015	1.000

Z_qs : standardized values of q^{th} variable

a_{kq}: element belonging to k^{th} eigenvector and for q^{th} variable, k=1,2, …,q; q=1,2, …,q.

Now, the composite index was constructed using the obtained eigenvalues of variables and principal components as under:

$$CI_i = \frac{\lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_q P_q}{\sum_{j=1}^{q}\lambda_j} \quad \ldots (3)$$

where,

CI_i = composite index for i^{th} district,

λ_js are eigen values,

P_q's are q^{th} principal components, i=1,2, …,n; j=1,2, …,q.

Further, the composite index of each district was normalized by using the following formula:

$$CI_{ni} = \frac{CI_i - \min(CI)}{\max(CI) - \min(CI)} \quad \ldots (4)$$

where,

Ci_{ni} = normalized value of composite index of i^{th} district,

min(CI)= minimum value of composite index among all,

max (CI) = maximum value of composite index among all.

Table 3a: Eigen vectors

	PC1	PC2	PC3	PC4
Rose	0.556	-0.027	-0.803	0.213
Gladiolus	0.122	0.991	0.056	0.021
Marigold	0.572	-0.114	0.557	0.591
Seasonal	0.591	-0.068	0.205	-0.778

Note: PC indicates the Principal Component.

Table 4: District wise Flower Production Index (FPI) along with their rank.

Districts	FPI	Rank
Purba Medinipur	1.000	1
Nadia	0.861	2
Paschim Medinipur	0.594	3
Darjeeling	0.419	4
North 24-Paraganas	0.250	5
Howrah	0.222	6
South 24-Paraganas	0.166	7
Jalpaiguri	0.027	8
Uttar Dinajpur	0.019	9
Burdwan	0.018	10
Murshidabad	0.016	11
Malda	0.011	12
Coochbehar	0.009	13
Bankura	0.009	14
Birbhum	0.007	15
Hooghly	0.002	16
Dakshin Dinajpur	0.002	17
Purulia	0.000	18

RESULTS AND DISCUSION

The analysis of collected data was done using SAS package (SAS Institute India Private Limited, Mumbai, India). The PRINCOMP procure was used to analyze the data. The obtained correlation matrix is given in table-1. The correlation between (PR, PM), (PR, PSF) and (PM, PSF) was found significant at p value 0.05. The correlation between remaining variable were not found significant at p value 0.05.

Also, regression analysis was performed and Variance Inflation Factor (VIF) for each variable was obtained to detect multicolinearity by regressing one variable to other remaining variables. The Variance Inflation Factor for jth variable can be obtained as under

$$VIF_j = \frac{1}{1 - R^2_j}$$

where,

VIF_j is Variance Inflation Factor for j^{th} variable.

Coefficients of determination (R_j^2) were obtained by regressing jth variable on other variable(s). The result of regression analysis along with VIFs is given in table 2. It was concluded that the linear relationship among variables is significant except for PG. It was also concluded that the variable PSF is having serious multicolinearity as VIF for PSF is greater than 10 (Montgomery *et al.,* 2001). Also the variable PM can be considered as near multicolinearity. While in case of PR there is very little multicolinearity and in case of PG, there is no multicolinearity. Overall, it was concluded that there is multicolinearity among variables. Thus, the composite index has been constructed using PCA to overcome the problem of multicolinearity.

The results of PRINCOMP procedure were obtained and the eigen values and eigenvectors are given in table 3 and 3a. It can be seen that only first two principal component accounts more than 90 % variability. The sensitivity of constructed composite index can also be verified by observing the components of eigenvectors. The indicators having highest component in first eigen vector influences maximum to the Composite Index (CI) that is the CI is highly sensitive to the associated indicators. Thus, it can be concluded that the constructed composite index is highly sensitive to production of Seasonal Flower followed by Marigold and Rose because the first principal component have highest value for Seasonal Flower followed by Marigold and Rose and also it has maximum eigen value 2.719.

The construction of Flower Production Index (FPI) was performed using the methodology discussed earlier. The composite index value for each district along with their rank is given in table 4 after arranging in descending order. The districts were identified and grouped in to two categories High and Low on the basis of constructed FPI. The FPI values greater than or equal to 75^{th} percentile were grouped as high flower production zone and the districts having FPI values less than 75^{th} percentile (Kumar *et al.,* 2013) were identified as low flower production zone. The 75^{th} percentile of FPI was found to be 0.35. Thus, Purba Medinipur, Nadia, Paschim Medinipur and Darjeeling were identified as high flower production zone whereas remaining districts were grouped as low flower production zone in West Bengal.

REFERENCES

Ahmad, T., Singh, R. and Rai, A. 2003. Development of GIS based technique for identification of potential agro forestry areas. *Proj. Rep.*, IASRI.

Dutta, P., Kundu, S., Bauri, F. K., Talang, H. and Majumder, D. 2014. Effect of bio-fertilizers on physico-chemical qualities and leaf mineral

composition of guava grown in alluvial zone of West Bengal. *J. Crop Weed,* **10**: 268-71.

Erkot, E. and Moran, S.R. 1991. Locating obnoxious facilities in public sector: An application of the analytic hierarchy process to the municipal landfill siting decisions. *Socio- Econ. Planing Sci.*, **25**: 89-102.

Kumar, M., Ahmad, T., Rai, A. and Sahoo, P. M. 2013. Methodology for construction of composite Index. *Int. J. Agril. Stat. Sci.,* **9**: 639-47.

Mahlberg, B. and Obersteiner, M. 2001. Remeasuring the HDI by data Envelopment analysis, *Interim Report* IR-01-069, International Institute for Applied System Analysis, Laxenburg, Austria.

Montgomery, D. C., Peck, E. A. and Vining, G. G. 2001. *Intro. to Linear Regression Anal.*, 3rd Ed., Wiley, New York.

Narain, Rai, S.C. and Sarup, S. 1991. Statistical evaluation of development on socio-economic front. *J. Indian Soc. Ag. Stat.,* **43**: 339-45

Narain, P., Sharma, S.D., Rai, S.C. and Bhatia, V.K. 2005. Dimension of socio-economic development in Jammu and Kashmir. *J. Indian Soc. Ag. Stat.,* **59**: 243-50.

Parker, J. 1991. Environmental reporting and environmental indices. *Ph.D. Dissertation*, Cambridge, U.K.

Siddique, M.Z., Everett, J.W., and Vieux, B.E. 1996. Landfill siting using geographical information system: A demonstration. *J. Env. Engg.*, **122**: 515-23.