

Original Research Article

<https://doi.org/10.20546/ijcmas.2017.604.311>

Cluster Analysis to Form Similarity for Major Selected Crops in Rajasthan, India

Lokesh Kumar Meena^{1*}, Chandra Sen² and Saket Kushwaha²

¹Department of Agricultural Economics, Bihar Agricultural University, Sabour-813210, Bhagalpur, Bihar (India)

²Department of Agricultural Economics, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi-221005 (India)

*Corresponding author:

ABSTRACT

The main objective of the study was to form similarity based on area, production and productivity for different crops of Rajasthan for which the secondary data on area, production and productivity of 23 different major crops i.e. period from 1980 to 2014 were collected. The study period was classified as Pre WTO era from 1980-1995 and Post WTO era from 1996-2014. Ward's method of clustering under hierarchical cluster analysis was used and Dendrogram shows the graphical representation of the results of hierarchical cluster analysis. Results found that wheat, mustard & rapeseed, gram cotton and bajra crop showed similarity in area across all the districts during the first period while in the second period only wheat, mustard & rapeseed and bajra crops had similarity in area across all the districts. Area of gram and cotton has shifted over the years in the second period of study. Results also indicated that wheat, gram, mustard & rapeseed, bajra and cotton had similarity in production across all the districts during the first period. While only wheat and bajra showed similarity in production across all the districts in second period. Wheat and bajra were the crops which made clusters in production across all the districts of Rajasthan from first period to second period. Potato and wheat crops had similarity in productivity across all the districts during the first period while the crops like coriander, potato, wheat, garlic and pea had similarity in productivity across all the districts of Rajasthan. Only wheat and bajra were the crops which had similarity in productivity across all the districts of Rajasthan from first period to second period.

Keywords

Cluster, Hierarchical Clustering, Ward Method, Euclidian Distance, Dendrogram.

Article Info

Accepted:
25 March 2017
Available Online:
10 April 2017

Introduction

Cluster analysis is a convenient method for identifying homogenous groups of objects called clusters. Objects (or cases, observations) in a specific cluster share many characteristics, but are very dissimilar to objects not belonging to that cluster. Now a day's Agricultural expert systems are being used extensively almost in every walk of life. Various tools also have been evolved for

evaluating, justifying, upgrading and modifying the existing agricultural expert systems thus making them more useful in their intended purposes⁸. Cluster analysis was used in this study. Its tool for exploratory data analysis and based on grouping of relatively homogenous classes^{7, 10}, the main aim of cluster analysis is an attempt to combine cases into groups when the group membership was

not known prior to the analysis. It is a technique for grouping individual or objects into unknown groups. In biology, cluster analysis has been used for decades in the area of taxonomy, where living things are classified into arbitrary groups on the basis of their characteristics group. The classification proceeds from the most general to the most specific in steps. The most general classification was kingdom followed by phylum, subphylum, and class etc⁶.

The Cluster analysis is a multivariate analysis which deals with grouping of objects or individuals into unknown objects based on the homogeneity of the objects. This method involves an agglomerative clustering algorithm. The procedure was starting out at the leaves and work its way to the trunk, so to speak. It looks for groups of leaves that it forms into branches, the branches into limbs and eventually into the trunk. Ward's method starts out with 'n' clusters of size one and continues until all the observations are included into one cluster⁹. The Dendrogram shows the graphical representation of the results of hierarchical cluster analysis. There was a tree-like plot where each step of hierarchical clustering was represented as a fusion of two branches of the tree into a single one. The branches represent clusters obtained on each step of hierarchical clustering^{2,4}.

Materials and Methods

The study was conducted across all districts of Rajasthan and based on the secondary data of area, production and productivity of 23 different crops period from 1980 to 2014 and collected from the Directorate of Economics and Statistics (DES), Jaipur and Institute of Social and Economic Change (ISEC), Bangalore. The study period was classified as Pre WTO era from 1980-1995 and Post WTO era from 1996-2014 to know the impact of agreement on Rajasthan agriculture cropping

pattern. The entire analytical frame work was carried out for both the periods indicated separates results for both the periods under study. Cluster analysis was used to form the similarities on area, production and productivity of different crops. The selected crops were gram, green gram, masur, pea, moth, cowpea, wheat, paddy, bajra, sorghum, maize, soybean, sesamum, ground nut, mustard and rapeseed, taramira, linseed, cotton, coriander, garlic, jeera or cumin, potato and methi.

Tools and techniques

Clustering methods

The cluster analysis is a multivariate procedure for segmentation applications in research. The statistical tools hierarchical cluster analysis was used, Ward's method of clustering under hierarchical cluster analysis was used to form similarity based on area, production and productivity for different crops of study area³.

Ward's method

It is an alternative approach for performing cluster analysis. Basically, it looks at cluster analysis as an analysis of variance problem, instead of using distance metrics or measures of association. This method involves an agglomerative clustering algorithm. It will start out at the leaves and work its way to the trunk, so to speak. It looks for groups of leaves that it forms into branches, the branches into limbs and eventually into the trunk. Ward's method starts out with *n* clusters of size 1 and continues until all the observations are included into one cluster^{6,7}.

This method is most appropriate technique for analyzing quantitative variables, and not for binary variables. Based on the notation that clusters of multivariate observations should

be approximately elliptical in shape, it is generally assumed that the data from each of the clusters will be realized in the pattern of multivariate distribution. Therefore, it follows elliptical shape when plotted in a p -dimensional scatter plot¹.

Notations used are as follows: X_{ijk} denote the value for variable 'k' in observation 'j' belonging to cluster 'i'. Further, under this method the following concepts are defined:

Error Sum of Square (ESS)

$$ESS = \sum_i \sum_j \sum_k (X_{ijk} - \bar{X}_{i.k})^2$$

Summing over all variables and all of the units within each cluster and comparing the individual observations for each variable against the cluster means for that variable. If the Error Sum of Squares is small, then it suggests that data are closer to their cluster means, implying that a cluster of like units⁶.

Total Sum of Squares (TSS)

$$TSS = \sum_i \sum_j \sum_k (X_{ijk} - \bar{X}_{.k})^2$$

The total sum of squares is defined in the similar way of ESS. Further, comparing the individual observations for each variable against the grand mean.

R-square (R^2)

$$R^2 = \frac{TSS - ESS}{TSS}$$

The R^2 value is interpreted as the proportion of variation explained by a particular clustering of the observations. Using Ward's Method indicate with all sample units in n clusters of size 1 each. In the first step of the algorithm, $(n-1)$ cluster are formed, one of

size two and the remaining of size one. The error sum of squares and R^2 values are then computed. The pair of sample units that yield the smallest error sum of squares, or equivalently, the largest R^2 value will form the first cluster. In the second step of the algorithm, $(n-2)$ clusters are formed from that $(n-1)$ clusters defined in step 2. This may include two clusters of size two or a single cluster of size three comprising of two items clustered in step 1. Again, the value of R^2 is maximized. Thus, at each step of the algorithm clusters or observations are combined in such a way to minimize the results of error from the squares or alternatively maximize the R^2 value. The algorithm stops when all sample units are combined into a single large cluster of size n .

Dendrogram

The Dendrogram is a graphical representation of the results of hierarchical cluster analysis. This appears in the form of tree like plot where, each step of hierarchical clustering is represented as a fusion of two branches of the tree resolving into a single one. The branches represent clusters obtained on each step of hierarchical clustering⁵.

Results and Discussion

Cluster Analysis

The cluster analysis was carried out based on area, production, and productivity of different agricultural and horticultural crops which were predominantly grown in the districts of Rajasthan and calculated for two different periods 1980-1995 and 1996-2014 independently.

Crop cluster based on area during 1980-1995

The cluster analysis was made based on the cultivated area of 23 different crops across all

the districts of Rajasthan for the first period of study i.e. 1980-1995 by measured Squared Euclidean Distance through Ward's method of clustering under hierarchical cluster analysis. From the obtained Dendrogram we can see that only three final major clusters were observed as given in Figure No. 1. The first cluster formed by wheat, mustard & rapeseed, gram and cotton. This indicates only these crops have similarity in area of cultivation across all the districts of Rajasthan. The second cluster mainly formed by bajra alone. This indicated that bajra crop has similarity in area of cultivation across all the districts of Rajasthan. Further, the last means third cluster was the combination of the rest of the crops viz; paddy, sorghum, maize, green gram, masur, pea, moth, cowpea, soybean, sesamum, ground nut, taramira, linseed, coriander, garlic, jeera or cumin potato and methi.

Crop cluster based on area during 1996-2014

The cluster analysis was made based on the cultivated area of 23 different crops across all the districts of Rajasthan for the first period of study i.e. 1996-2014 by measured Squared Euclidean Distance through Ward's method of clustering under hierarchical cluster analysis. From the obtained Dendrogram we can see that only three final major clusters were observed as given in Figure No.2. The first cluster formed by wheat and mustard & rapeseed. This indicated that these crops have similarity in area of cultivation across all the districts of Rajasthan. The second cluster mainly formed by bajra. This indicated that this crop have similarity in area of cultivation across all the districts of Rajasthan. Third cluster formed by remaining other crops paddy, sorghum, maize, green gram, masur, pea, moth, cowpea, soybean, sesamum, ground nut, taramira, linseed, coriander, garlic, jeera or cumin potato, methi, gram and

cotton. Based on this Dendrogram, it was evident that the area of cotton and gram has been shifted over the years during the period 1980-1995, when compared to the area during 1996-2014.

Crop cluster based on production during 1980-1995

Dendrogram (Figure No.3) showed that the cluster analysis done based on the Production of different crops across all the districts of Rajasthan for the period 1980-1995 by used Euclidean Distance through Ward's method of clustering under hierarchical cluster analysis. The first cluster was formed by wheat alone. Indicated that this crop has similar pattern of production across all the district of Rajasthan, the second cluster was mainly formed by gram, mustard and rapeseed, bajra and cotton. This indicated that these crops showed similarity in its production across all the districts of Rajasthan. Further the third cluster was formed with remaining crops paddy, sorghum, maize, green gram, masur, pea, moth, cowpea, soybean, sesamum, ground nut, taramira, linseed, coriander, garlic, jeera or cumin potato, and methi, indicated these crops have similarity in production across all the districts of Rajasthan.

Crop cluster based on production during 1996-2014

The cluster analysis was made based on production of 23 different crops across all the districts of Rajasthan for the second period of study i.e. 1996-2014 by measured squared Euclidean distance through ward's method of clustering under hierarchical cluster analysis. From the obtained dendrogram we can see that only three final major clusters were observed as given in Figure No. 4. From the Dendrogram we can see that the first cluster was formed by wheat only. This indicates that

wheat has similarity in Production across all the districts of Rajasthan. The second cluster was formed by bajra, indicated that this crop has similarity in Production across all the districts of Rajasthan. And final means third cluster was formed with all other remaining crops 21 crops viz; gram, mustard and rapeseed, cotton, paddy, sorghum, maize, green gram, masur, pea, moth, cowpea, soybean, sesamum, ground nut, taramira, linseed, coriander, garlic, jeera or cumin potato, and methi, showed similarity in production pattern across all the districts of Rajasthan. The Dendrogram showed that the gram, mustard and rapeseed and cotton crops were shifted, because these crops were undertaking cluster during the first period 1980-1995. But it was not formed any cluster during the period 1996-2014.

Crop cluster based on productivity or yield during 1980-1995

The cluster analysis was made based on the yield of 23 different crops across all the districts of Rajasthan for the first period of study i.e. 1980-1995 by measured Squared Euclidean Distance through Ward's method of clustering under hierarchical cluster analysis. From the obtained Dendrogram we can see that only three final major clusters were observed as given in Figure No.5. Three final clusters formed, the first cluster formed by potato crop. This indicated that potato alone showed similarity in productivity across all the districts of Rajasthan. The second cluster was formed by wheat. This indicates that this crop have similarity in productivity across all the districts of Rajasthan and third cluster formed by remaining other crops viz; gram, mustard and rapeseed, bajra, cotton, paddy, sorghum, maize, green gram, masur, pea, moth, cowpea, soybean, sesamum, ground nut, taramira, linseed, coriander, garlic, jeera or cumin, and methi. Based on the cluster Dendrogram, it can be concluded

that potato and wheat crops were introduced as a major crop across all the districts showed similarity in productivity during the period 1980-1995.

Cluster based crop productivity during 1996-2014

The cluster analysis was made based on the yield of 23 different crops across all the districts of Rajasthan for the second period of study i.e. 1980-1995 by measured Squared Euclidean Distance through Ward's method of clustering under hierarchical cluster analysis. From the obtained Dendrogram we can see that only three final major clusters were observed as given in Figure No.6. Three final clusters were formed; the first cluster was formed by coriander and potato indicated similarity in productivity of these crops across all the districts of Rajasthan. The second cluster was formed by wheat, garlic and pea. This indicated these were the only crops which showed similarity in productivity or yield across all the districts of Rajasthan. The final third cluster was formed by considering remaining other crops viz; gram, mustard and rapeseed, bajra, cotton, paddy, sorghum, maize, green gram, masur, moth, cowpea, soybean, sesamum, ground nut, taramira, linseed, jeera or cumin, and methi. This indicated that remaining other crops have similarity in productivity of crops across all the districts of Rajasthan. Based on the cluster Dendrogram, it can be concluded that coriander, garlic and pea crops were introduced as a major crop across all the districts showed similarity in productivity during the period 1996-2014.

Overall performance of the cluster analysis

For first period of crop area between 1980-1995, various crops such as wheat, mustard & rapeseed, gram cotton and bajra showed similarity in area across all the districts in

Rajasthan. For the second period of the study i.e. 1996-2014, only wheat, mustard & rapeseed and bajra crops had similarity in area across all the districts of Rajasthan. When the performance of crop clusters based on area between two periods was compared, it was evident that gram and cotton has shifted over the years in the second period of study because both the crops were made clusters in the first period but not in the second period. Results also showed that for the period of 1980-1995, wheat, gram, mustard & rapeseed, bajra and cotton had similarity in production across all the districts of Rajasthan. For the second period of 1996-2014, only wheat and bajra showed similarity in production across all the districts of Rajasthan. When comparison of the performances of crop clusters based on production between two periods was observed that gram, mustard & rapeseed and cotton production shifted over the period it means these crops were made cluster in the first period but not in the second period. While wheat and bajra were the crops which made clusters or had similarity in production across all the districts of Rajasthan from first period to second period.

Table No. 1 also indicated that for the first period (1980-1995) of the study potato and wheat crops had similarity in productivity across all the districts of Rajasthan. For

second period 1996-2014, the crops like coriander, potato, wheat, garlic and pea showed similarity or made clusters in productivity across all the districts of Rajasthan. When the performances of crop clusters were compared based on productivity between two periods, it indicated that coriander, garlic and pea productivity included over the years in the second period of the study it means these crops were made clusters in the second period but not in the first period. Only wheat and bajra were the crops which had similarity in productivity across all the districts of Rajasthan from first period to second period.

In conclusion, from the present study we concluded that when the performance of crop clusters based on area between two periods was compared, it was evident that gram and cotton has shifted over the years in the second period of study. When comparison of the performances of crop clusters based on production between two periods was observed that gram, mustard & rapeseed and cotton production shifted over the period. It means these crops were made cluster in the first period but not in the second period. While wheat and bajra were the crops which made clusters or had similarity in production across all the districts of Rajasthan from first period to second period.

Table.1 Overall performance of the cluster analysis

Aspects	Periods	First cluster	Second cluster	Third cluster
Area	1980-1995	Wheat, Mustard & rapeseed, Gram and Cotton (04)	Bajra (01)	Other remaining crops (18)
	1996-2014	Wheat and Mustard & rapeseed (02)	Bajra (01)	Other remaining crops (20)
Production	1980-1995	Wheat (01)	Gram, Mustard & Rapeseed, Bajra and Cotton. (04)	Other remaining crops (18)
	1996-2014	Wheat (01)	Bajra (01)	Other remaining crops (21)
Productivity	1980-1995	Potato (01)	Wheat (01)	Other remaining crops (21)
	1996-2014	Coriander and Potato (02)	Wheat, Garlic and Pea (03)	Other remaining crops (18)

Fig.1 Crop cluster based on area during 1980-1995

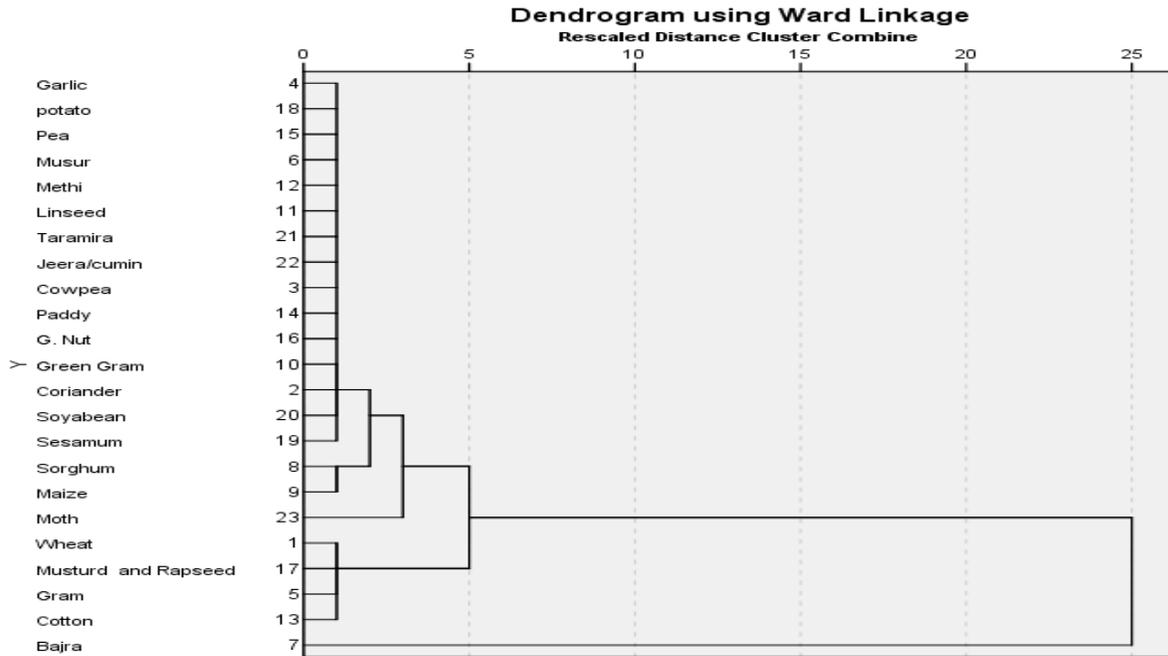


Fig.2 Crop cluster based on area during 1996-2014

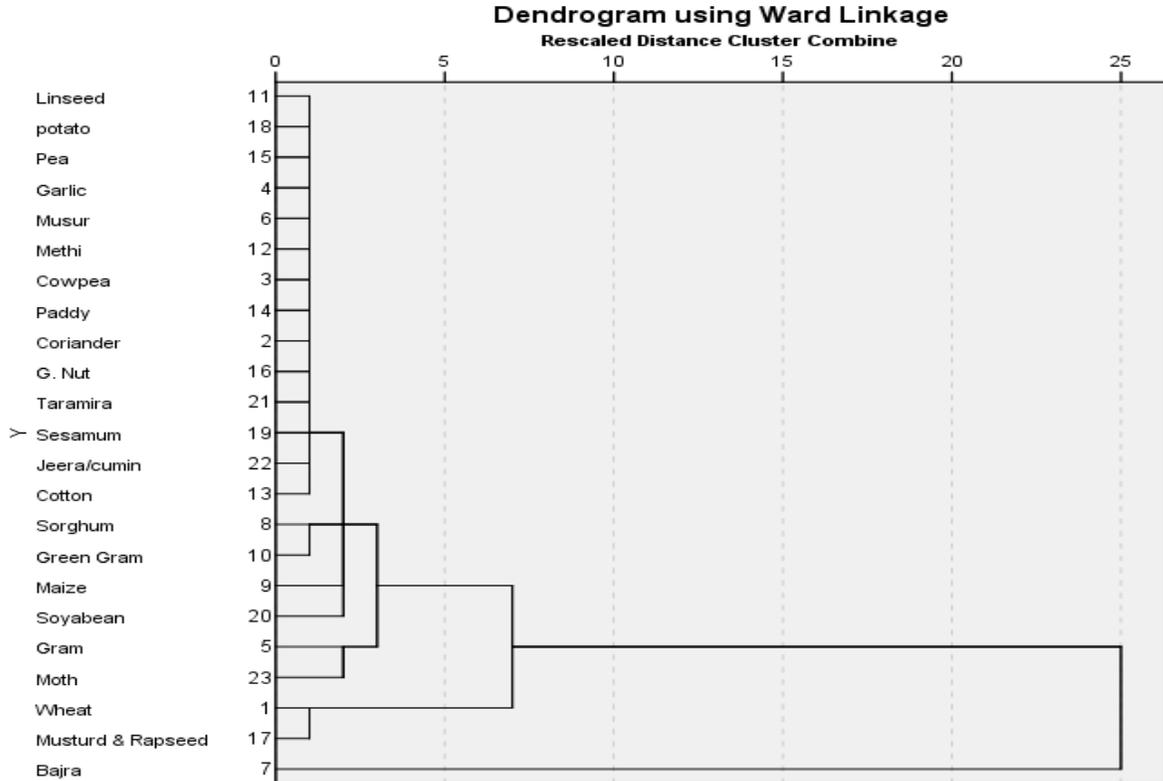


Fig.3 Crop cluster based on production during 1980-1995

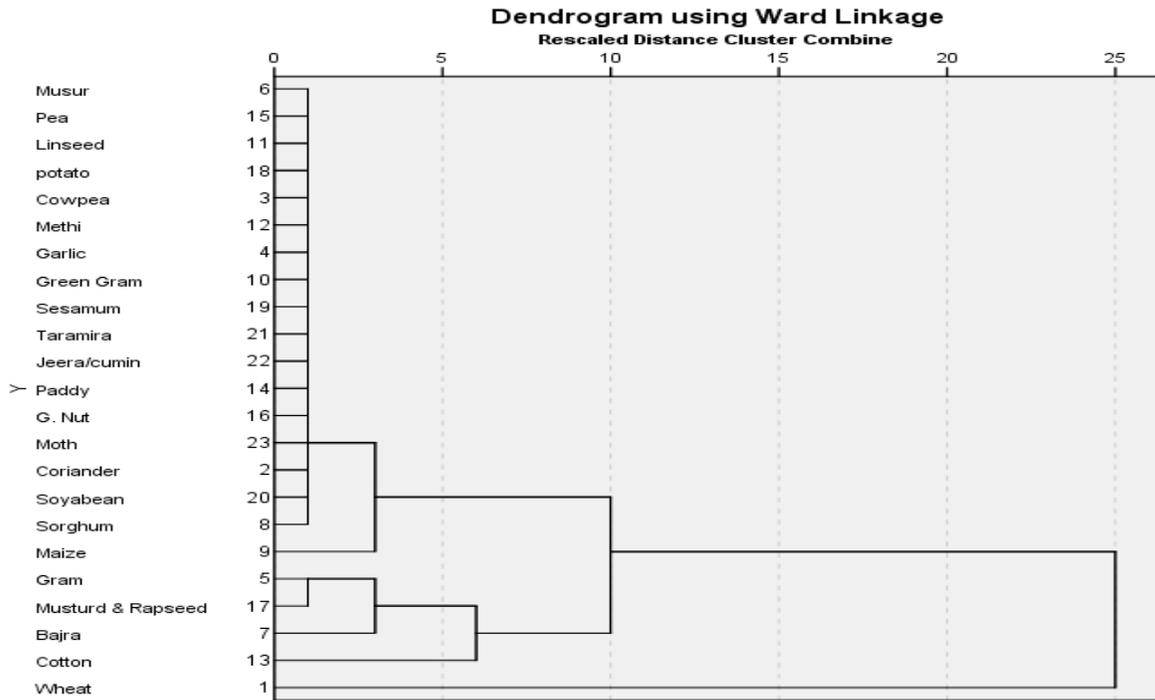


Fig.4 Crop cluster based on production during 1996-2014

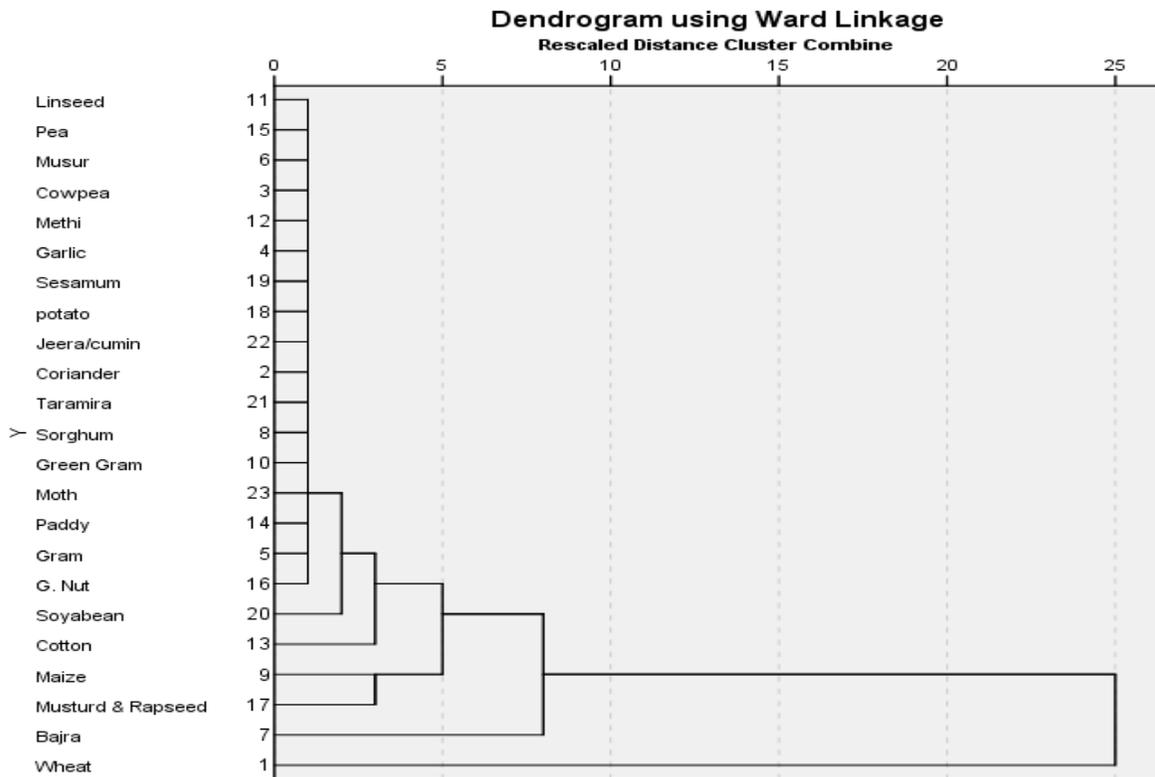


Fig.5 Crop cluster based on productivity during 1980-1995

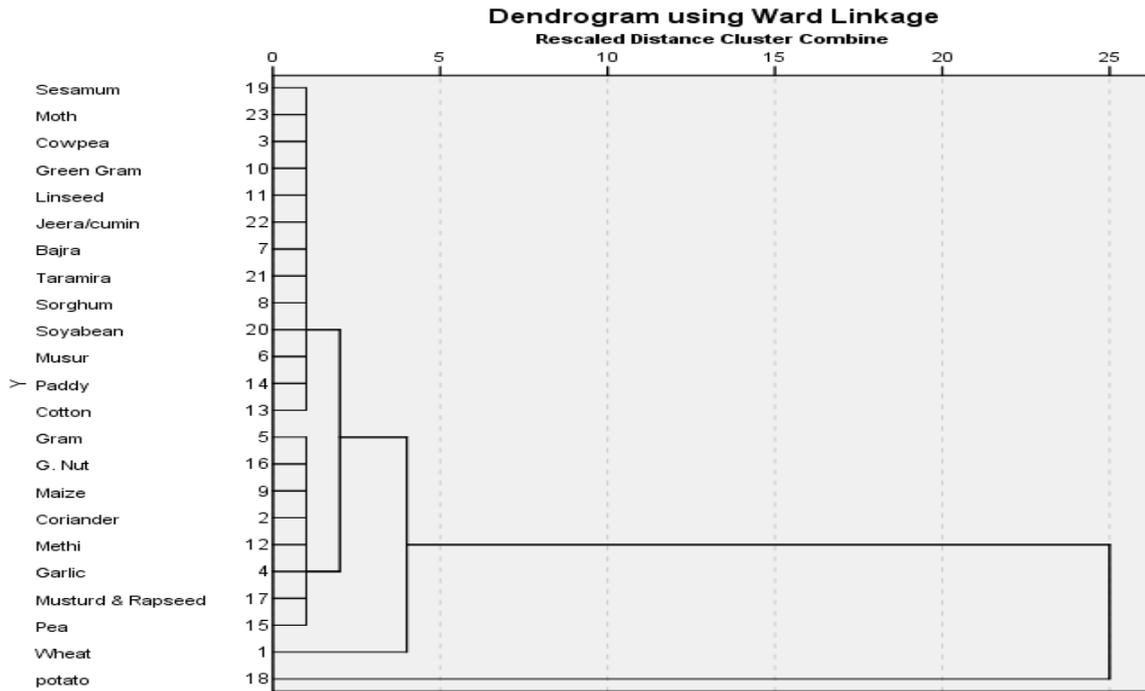
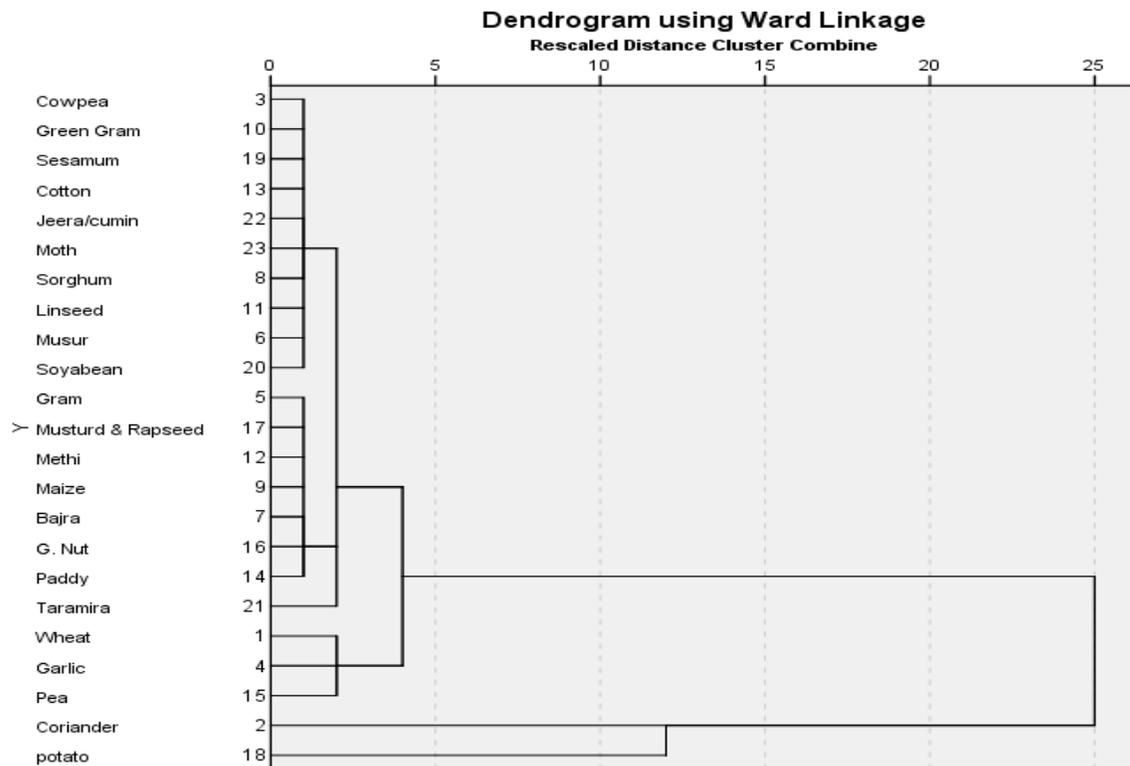


Fig.6 Cluster based crop productivity during 1996-2014



The present study also concluded that horticultural crops had similarity in productivity across all the districts of Rajasthan during the both period. It means coriander, garlic and pea productivity included over the years in the second period of the study. Only wheat and bajra were the crops which had similarity in productivity across all the districts of Rajasthan from first period to second period.

Acknowledgement

It is my Ph.D work conducted in across all districts of Rajasthan at Institute of Agricultural Sciences, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi-221055 (India), under the guidance of Professor Chandra Sen, Professor, Department of Agricultural Economics, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi-221005 (India). I am really thankful to my advisor for giving his guidance time to time.

References

1. Batagelj, V., Classification and Related Methods of Data Analysis. H.H. Bock (editor). North- Holland, Amsterdam, 1988. p. 67-74
2. Cluster analysis” in <http://public.fh-wolfenbuettel.de/~hoeppnef/clustering.html>
3. Guptha, R.P. and Tewari, S.K., Examined that the Factors Affecting Crop Diversification: An Empirical Analysis, *Indian Journal of Agril. Economics*, 2004, 45 (3), 304-307.
4. Gupta A, Gupta A and Mishra A, Research Paper on Cluster Techniques of data Variations. *International Journal of Advance Technology & Engineering Research*, 2011 1: 39-47.
5. Han, J., Kamber, and M., Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann, September 2000).
6. https://en.wikipedia.org/wiki/Cluster_analysis
7. https://en.wikipedia.org/wiki/Ward%27s_method
8. Johnson, R.A. and Wichern, D.W., Applied Multivariate Statistical Analysis, PHI publication New Delhi, V edn, 2000, 668-748.
9. Kaufman. L and Rousseeuw, P.J., Finding groups in data: an application to cluster analysis, Wiley, New York. 2005.
10. Kumari, M. and Misra, B., Application of Cluster Analysis in Agriculture A Review Article, *International Journal of Computer Applications*, 2011, 36 (4), 43-47.
11. Rathod, S., Surendra, H.S., Munirajappa and Chandrasekhar H., Cluster Analysis on Area, Production and Productivity for Major Selected Crops in Karnataka. *Mysore J. Agric. Sci.*, 2012, 46 (2): 293-299.
12. Wiesek, A.K., use of cluster analysis in classification similarities in variable associated with agricultural green house gases emissions in OECD country, 2015.
13. Ward, J. H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 1963, 58, 236–244.

How to cite this article:

Lokesh Kumar Meena, Chandra Sen and Saket Kushwaha. 2017. Cluster Analysis to Form Similarity for Major Selected Crops in Rajasthan. *Int.J.Curr.Microbiol.App.Sci.* 6(4): 2673-2682. doi: <https://doi.org/10.20546/ijcmas.2017.604.011>