

Cluster Analysis for Agriculture and other allied sectors

Dr. B.Sarojamma¹, K.Geetha², Dr. B. HariMallikarjuna Reddy³

¹Assistant Professor, Department of Statistics,
S.V.University, Tirupati.

²Research scholar S.V.University, Tirupati.

³Academic consultant, Department of Statistics,
S.V.University, Tirupati.

ABSTRACT

Agricultural Sector is the major and popular source in India. Agriculture and allied sectors are considered to be main stream of the Indian Economy. So many Industries are depending on Agricultural Sector. Some of them are fertilizers industry, pesticides industry etc. Agriculture and allied industry is divided into several segments namely Horticulture sector, Fisheries sector, Sericulture Sector etc. Our country is the second largest producer of fruits and vegetables in the world. In this paper, we are taken industries like Agriculture, crops, Livestock, Afforestation, Acquafarming, Mining and Quarrying, Manufacturing, Electricity and gas, construction and other services for 2011-12 to 2017-18. By using Factor analysis, we are dividing them into clusters and Dendrograms are drawn appropriately.

Keywords: Agriculture and allied sectors, Cluster Analysis, Dendrograms.

1. INTRODUCTION

Agriculture plays a major role in India. Today itself many people are living on Agriculture sector. Agriculture means not only paddy yield but also pulses, cereals etc. agriculture is the back bone of India. India is the second largest Food Producer as per 2018. Agriculture employed 50% of Indian work force and contributed 17-18% to Country's Gross Domestic Product. "Economic Development and Agriculture in India" by Adam Cagliari and Anthony Rush explains key development in the Agriculture Sector in India including productivity gains and integration with global markets.

2. METHODOLOGY

For Factorization of Agriculture and Allied Sectors mentioned above, we applied concept of cluster analysis.

CLUSTER ANALYSIS

Cluster analysis is nothing but grouping of variables. Each cluster is having the similar characteristic variable for solving cluster analysis. In fact, there are more than 100 clustering algorithm are there. A few among them are:

- Hierarchical Clustering
- Centroid based clustering
- Distribution based clustering

- Density based clustering etc.

Among all, we chose Hierarchical Cluster Analysis which follows steps as Calculating Distances, Link the Distances and choosing a solution by selecting right number of cluster.

SPSS PROCEDURE FOR ESTIMATION CLUSTER ANALYSIS

- Step1. Enter variables i.e. in variable view
- Step2. Data according to variables in Data view.
- Step3. Go to “Analyze” in menu bar
- Step4. In Analyze go to “Classify”
- Step5. In that go to “Hierarchical cluster”
- Step6. Send all variables to “Variables” menu
- Step7. Go to “Statistics” in that take “Proximity matrix” click on continue
- Step8. Go to plots in that tick “Dendrogram” vertical or horizontal orientation which ever you want.
- Step9. Go to “Method” in that select “Between groups linkage” and interval measure method “Squared Euclidean distance”
- Step10. Cluster according to variables
- Step11. In display tick “Statistics” and “Plots”.
- Step12. Click “Ok”.

DENDROGRAM

Dendrogram is a visual presentation that is frequently used in various fields like Hierarchical clustering, biology, Genetics etc.

3. EMPIRICAL INVESTIGATION

By using hierarchical clustering performed in SPSS as

- Analyze → Classify → Hierarchical clustering → Variables → 1. Statistics 2. Plots 3. Method.
- Statistics → Proximity matrix → Continue.
- Plots → Dendrogram → All Clusters → Vertical → Continue
- Method → 1. Cluster Method → between groups linkage 2. Interval → Squared Euclidean distance → Continue
- Finally click ‘Ok’

THE PROXIMITY MATRIX OF CONSIDERED VARIABLES IS GIVEN BELOW

CASES	MATRIX FILE INPUT										
	AGRICULTURE FORESTRY	CROPS	LIVE STOCK	FORESTRY AND LOGGING	FISHING AND AQUACULTURE	INDUSTRY	MINING QUARRYING	MANUFACTURING	ELECTRICITY AND GAS	CONSTRUCTION	SERVICES
Agriculture Forestry	.000	296.860	1314.840	1923.680	2042.210	1089.410	1636.340	12.620	1690.290	648.770	7877.740
crops	296.860	.000	366.780	714.000	786.790	2515.010	543.080	195.520	575.290	71.770	11209.960
livestock	1314.840	366.780	.000	59.060	80.530	4793.830	20.300	1088.220	23.970	124.270	15588.300
Forestry and logging	1923.680	714.000	59.060	.000	1.890	5902.010	12.080	1648.640	7.850	342.830	17558.900
Fishing and aquaculture	2042.210	786.790	80.530	1.890	.000	6109.040	23.370	1758.090	16.720	394.840	17908.110
Industry	1089.410	2515.010	4793.830	5902.010	6109.040	.000	5387.810	1315.290	5488.980	3407.140	3153.350
Mining quarrying	1636.340	543.080	20.300	12.080	23.370	5387.810	.000	1383.920	1.790	227.070	16675.480
manufacturing	12.620	195.520	1088.220	1648.640	1758.090	1315.290	1383.920	.000	1432.750	495.910	8458.360
ElectricityAnd gas	1690.290	575.290	23.970	7.850	16.720	5488.980	1.790	1432.750	.000	250.980	16832.050
construction	648.770	71.770	124.270	342.830	394.840	3407.140	227.070	495.910	250.980	.000	13043.350
Services	7877.740	11209.960	15588.300	17558.900	17908.110	3153.350	16675.480	8458.360	16832.050	13043.350	.000

TABLE NO 1

Generally proximity matrix is a squared matrix which gives the measurement of how extent similarities and dissimilarities occurs.

The above matrix gives shortest value as 1.790 for “Electricity and Gas” and “Minimum Quarrying” and the last distance value is between “services” and “construction” distance is 13043.350.

AGGLOMERATION SCHEDULE

STAGE	CLUSTER COMBINED		COEFFICIENTS	STAGE CLUSTER FIRST APPEARS		NEXT STAGE
	CLUSTER 1	CLUSTER 2		CLUSTER 1	CLUSTER 2	
1	7	9	1.790	0	0	4
2	4	5	1.890	0	0	4
3	1	8	12.620	0	0	7
4	4	7	15.005	2	1	5
5	3	4	45.965	0	4	8
6	2	10	71.770	0	0	7
7	1	2	409.265	3	6	8
8	1	3	1012.246	7	5	10
9	6	11	3153.350	0	0	10
10	1	6	8953.376	8	9	0

TABLE NO 2

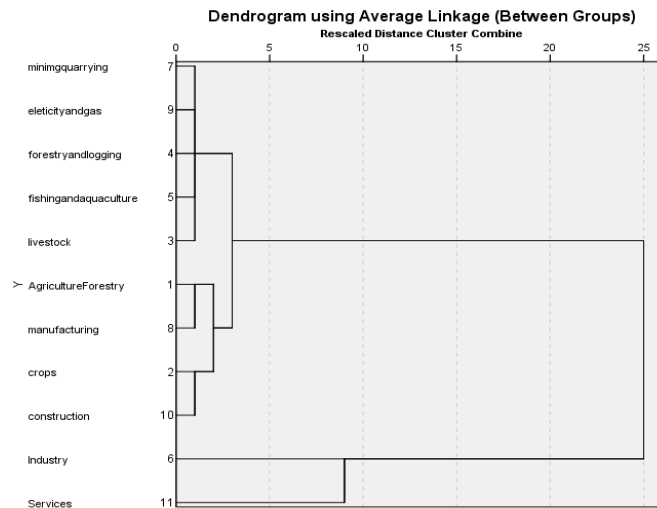


FIG NO 1

Upon watching Dendrogram, We had

1. Mining & quarrying 2. Electricity, gas 3. Forestry and logging 4. Fishing and aquaculture 5. Livestock as one cluster and another cluster is 1. Agriculture, Forestry 2. Manufacturing 3. Crops 4. Construction and third group is 1. Industry and 2. Services

The distance between first cluster and second cluster is distance is 2.5 and these two clusters with third cluster are 15.

4. SUMMARY AND CONCLUSIONS

All the eleven components are clustered into three groups

Cluster1: 1. Mining & quarrying 2. Electricity, gas 3. Forestry and logging 4. Fishing and aquaculture 5. Live stock

Cluster2: 1. Agriculture, Forestry 2. Manufacturing 3. Crops 4. Construction

Cluster3: 1. Industry and 2. Services. The dendrogram has drawn for various variables.

5. REFERENCE

- [1] Adam Cagliarini and Anthony Rush (June 2011). "Bulletin: Economic Development and Agriculture in India", Reserve Bank Of Australia PP.15-22.
- [2] L.P. Yuan (2010) "A Scientific's perspective on Experience with SRI in China for Raising the yields of Super Hybrid Rice". Acheied from original on Nov 2011.