Quantitative Methods for SOCIAL SCIENCES

Edited by: Vinayak Nikam • Abimanyu Jhajhria• Suresh Pal

This reference book is designed keeping in mind the need for the application of advanced quantitative methods in social science research to enhance its accuracy. The chapters are written in such a way that social scientists can easily grasp the methods including their theoretical and practical aspects using statistical software. The book provides comprehensive coverage of multivariate techniques, forecasting methods, structural equations, optimization models, quantitative methods for impact assessment, growth models and other important methods used in social science research.

ABOUT THE EDITORS

Vinayak Nikam is working as Scientist (Senior Scale) at ICAR-National Institute of Agricultural Economics and Policy Research (New Delhi). He did his PhD from ICAR-Indian Agricultural Research Institute (New Delhi) in agricultural extension. He also served ICAR-Central Soil Salinity Research Institute (Karnal) before joining NIAP in Nov 2015. He also holds faculty membership in the discipline of Agricultural Extension at ICAR-Indian Agriculture Research Institute (New Delhi). Currently, he is working on the performance and impact assessment of Agriculture Extension and Advisory Services as well as associated with technology impact assessment.

Abimanyu Jhajhria is a Scientist (Agricultural Economics) at ICAR-National Institute of Agricultural Economics and Policy Research (New Delhi). He has received his doctorate in Agricultural Economics from the ICAR-Indian Agricultural Research Institute (New Delhi). His research interest includes markets, trade and institutions. He is involved in the projects on market reforms, agricultural commodity value chains and outlook models for agricultural commodities.

Suresh Pal is Director of ICAR-National Institute of Agricultural Economics and Policy Research (New Delhi). He has a PhD in agricultural economics from ICAR-Indian Agricultural Research Institute (New Delhi). He has published extensively on different aspects of Indian agriculture and guided doctoral and masters students. He has received awards for his contributions like best journal article awards, Norman E Borlaug International Science Fellowship, Fellow of the Indian Society of Agricultural Economics (Mumbai), and Fellow of the National Academy of Agricultural Sciences.



ICAR-NATIONAL INSTITUTE OF AGRICULTURAL ECONOMICS AND POLICY RESEARCH (NIAP) D.P.S.Marg, Pusa, New Delhi – 110012. Phone: +91-11- 25847628, +91-11- 25848731 Fax: +91-11-25842684 . E-mail:director.niap@icar.gov.in; Website: www.http://www.ncap.res.in

SS S Quantitative Methods for **Social Sciences** Vinayak Nikam • Abimanyu Jhajhria • Suresh Pa

Quantitative Methods for SOCIAL SCIENCES



ICAR-National Institute of Agricultural Economics and Policy Research

Reference book on

Quantitative Methods for **SOCIAL SCIENCES**

Edited by

Vinayak Nikam Abimanyu Jhajhria Suresh Pal



ICAR-National Institute of Agricultural Economics and Policy Research New Delhi

Quantitative Methods for **SOCIAL SCIENCES**

Edited by Vinayak Nikam, Abimanyu Jhajhria and Suresh Pal

© 2019 ICAR-National Institute of Agricultural Economics and Policy Research

Published by Dr. Suresh Pal Director, ICAR-National Institute of Agricultural Economics and Policy Research

ISBN: 978-81-940080-2-6

Printed at Chandu Press: chandupress@gmail.com

CONTENTS

| Prefe | ace | 111 |
|-------|--|-----|
| Ackr | nowledgements | v |
| 1. | Overview of quantitative methods for social science research Vinayak Nikam, Abimanyu Jhajhria and Suresh Pal | 1 |
| Part | I: Measures of interdependence of variables/cases | |
| 2. | Cluster analysis Arpan Bhowmik, Sukanta Dash, Seema Jaggi and Sujit Sarkar | 7 |
| 3. | Principal component analysis Prem Chand, M. S. Raman and Vinita Kanwal | 19 |
| 4. | Multidimensional scaling Ramasubramanian V. | 31 |
| 5. | Correspondence analysis Deepak Singh, Raju Kumar, Ankur Biswas, R. S. Shekhawat and Abimanyu Jhajhria | 46 |
| Part | II: Regression analysis | |
| 6. | Linear and non-linear regression analysis Ranjit Kumar Paul and L. M. Bhar | 59 |
| 7. | Qualitative regression model (Logit, Probit, Tobit) Shivaswamy G. P., K. N. Singh and Anuja A. R. | 70 |
| 8. | Introduction to panel data regression models Ravindra Singh Shekhawat, K. N. Singh, Achal Lama and Bishal Gurung | 78 |
| 9. | Auto regressive and distributed lag models Rajesh T., Harish Kumar H. V., Anuja A. R. and Shivaswamy G. P. | 88 |
| 10. | Conjoint analysis Sukanta Dash, Krishan Lal and Rajender Parsad | 96 |
| 11. | Two stage least square simultaneous equation model Shivendra Kumar Srivastava and Jaspal Singh | 110 |

| 12. | Discriminant function analysis Achal Lama, K. N. Singh, R. S. Shekhawat, Kanchan Sinha and Bishal Gurung | 121 |
|------|---|-----|
| Part | III: Time series analysis | |
| 13. | Price forecasting using ARIMA model Raka Saxena, Ranjit Kumar Paul and Rohit Kumar | 129 |
| 14. | Volatility models Girish Kumar Jha and Achal Lama | 142 |
| 15. | Artificial neural network for time series modelling Mrinmoy Ray, K. N. Singh, Kanchan Sinha and Shivaswamy G. P. | 155 |
| 16. | Hybrid time series models Ranjeet Kumar Paul | 163 |
| Part | IV: Impact assessment methods | |
| 17. | Economic surplus approach Vinayak Nikam, Jaiprakash Bishen, T. K. Immanuelraj, Shiv Kumar and Abimanyu Jhajhria | 177 |
| 18. | Introduction to causal inference Arathy Ashok | 192 |
| 19. | Propensity score matching K. S. Aditya and Subash S. P. | 199 |
| 20. | Difference-in-difference model M. Balasubramanian and Gourav Kumar Vani | 211 |
| 21. | Regression discontinuity design Subash S. P. and Aditya K. S. | 219 |
| 22. | Synthetic control method Prabhat Kishore | 230 |
| 23. | Instrumental variable estimation Anuja A. R., K. N. Singh, Shivaswamy G. P., Rajesh T. and Harish Kumar H. V. | 236 |
| Part | V: Growth analysis | |
| 24. | Computable general equilibrium models Balaji S. J. | 245 |

| 25. | Decomposition of total factor productivity: DEA approach Dharam Raj Singh, Suresh Kumar, Venkatesh P. and Philip Kuriachen | 254 |
|------|--|-----|
| 26. | Total factor productivity using stochastic production function Shiv Kumar, Abdulla and Deepak Singh | 264 |
| Part | VI: Other methods | |
| 27. | Linear programming Harish Kumar H. V., Rajesh T., Shivaswamy G. P. and Anuja A. R. | 277 |
| 28. | Multi objective programming Chandra Sen | 289 |
| 29. | Structural equation modelling P. Sethuraman Sivakumar, N. Sivaramane and P. Adhiguru | 296 |
| 30. | Partial equilibrium model Shinoj Parappurathu | 310 |
| 31. | Production function analysis Suresh Kumar, Dharam Raj Singh and Girish Kumar Jha | 319 |
| 32. | Social network analysis Subash S. P. | 335 |
| 33. | Construction of composite index Prem Chand | 351 |
| 34. | Basic scaling techniques in social sciences Sudipta Paul | 361 |
| 35. | Analytical hierarchy process: A multi-criteria decision making technique Anirban Mukherjee, Mrinmoy Ray and Kumari Shubha | 371 |
| 36. | Artificial intelligence, machine learning and big data Rajni Jain, Shabana Begam, Sapna Nigam and Vaijunath | 378 |
| | List of contributors | 395 |

Chapter 2

CLUSTER ANALYSIS

Arpan Bhowmik, Sukanta Dash, Seema Jaggi and Sujit Sarkar

INTRODUCTION

Statistical science plays a major role in any scientific investigation. Use of appropriate statistical techniques for analyzing the data is very crucial to obtain a meaningful interpretation of the investigation. Throughout any scientific inquiry which is an iterative learning process, variables are often added or deleted from the study. Thus, the complexities of most phenomena require an investigator to collect observations on many different variables which leads to the study of multivariate analysis.

Cluster analysis is an important statistical tool with respect to multivariate exploratory data analysis. It involves intricate techniques, methods and algorithms which can be applied in various fields, including economics and other social research. The aim of cluster analysis is to identify groups of similar objects (e.g. countries, enterprises, households) according to selected variables (e.g. unemployment rate of men and women in different countries, deprivation indicators of households, etc.). Cluster analysis is typically used in the exploratory phase of research when the researcher does not have any pre-conceived hypotheses or prior knowledge regarding the similarity of the objects. It is commonly not only the statistical method used, but rather is done in the early stages of a project to help guide the rest of the analysis (Timm, 2002, Hair *et al.*, 2006).

Cluster analysis differs from other methods of classification such as Discriminant analysis where classification pertains to known number of groups and the operational objective is to assign new observations to one of these groups. Whereas cluster analysis is a more primitive tool as in that no assumptions are made about the number of groups or the group structure and the grouping is done based on similarities or distances (dissimilarities).

Cluster analysis is also an important tool for investigation in data mining. For example, in marketing research consumers can be grouped on the basis of their preferences. In short it is possible to find application of cluster analysis in any field of research.

CLUSTERING METHODS

The commonly used methods of clustering are divided into two categories (Johnson and Wichern, 2006).

- (i) Hierarchical and
- (ii) Non-Hierarchical.

Hierarchical Cluster Analysis

Hierarchical clustering techniques proceed by either a series of mergers or a series of successive divisions. Agglomerative hierarchical method starts with the individent objects, thus there are as many clusters as objects. The most similar objects are for a series of their similarities. Eventual as the similarity decreases, all sub groups are fused into a single cluster.

Divisive hierarchical methods work in the opposite direction. An initial single go of objects is divided into two sub groups such that the objects in one sub group are from the objects in the others sub groups. These sub groups are then further divide into dissimilar sub groups. The process continues until there are as many sub group as objects i.e., until each object form a group. The results of both agglomerative and divisive method may be displayed in the form of two dimensional diagram knows Dendrogram. It can be seen that the Dendrogram illustrate the mergers or divisions the have been made at successive levels.

Linkage methods are suitable for clustering items, as well as variables. This is m true for all hierarchical agglomerative procedures. The following linkage are m discussed:

- (i) single linkage (minimum distance or nearest neighbour)
- (ii) complete linkage (maximum distance or farthest neighbour)
- (iii) average linkage (average distances)

Other methods of hierarchical clustering techniques like Ward's method and Centrol method are also available in literature.

Steps of agglomeration in Hierarchical cluster analysis

The following are the steps in the agglomerative hierarchical clustering algorithm for groups of N objects (items or variables).

- i. Start with N clusters, each containing a single entity and an N×N symmetric matrix of distance (or similarities) $D = \{dik\}$.
- ii. Search the distance matrix for the nearest (most similar) pair of clusters. Let the distance between most similar clusters U and V be duv.
- iii. Merge clusters U and V. Label the newly formed cluster (UV). Update the entries in the distance matrix by (a) deleting the rows and column giving the distances between cluster (UV) and the remaining clusters.

iv. Repeat steps (ii) and (iii) a total of N-1 times (all objects will be in a single cluster after the algorithm terminates). Record the identity of clusters that are merged and the levels (distances or similarities) at which the mergers take place.

The basic ideas behind the cluster analysis are now shown by presenting the algorithm components of linkage methods.

Non Hierarchical Clustering Method

Non Hierarchical clustering techniques are designed to group items, rather than variables, into a collection of K clusters. The number of clusters, K, may either be specified in advance or determined as part of the clustering procedure. Because a matrix of distance does not have to be determined and the basic data do not have to be stored during the computer run. Non hierarchical methods can be applied to much larger data sets than can hierarchical techniques. Non hierarchical methods start from either (1) an initial partition of items into groups or (2) an initial set of seed points which will form nuclei of the cluster.

K Means Clustering

The K means clustering is a popular non hierarchical clustering technique. For a specified number of clusters K the basic algorithm proceeds in the following steps (Afifi, Clark and Marg, 2004).

- i. Divide the data into K initial cluster. The number of these clusters may be specified by the user or may be selected by the program according to an arbitrary procedure.
- ii. Calculate the means or centroid of the K clusters.
- iii. For a given case, calculate its distance to each centroid. If the case is closest to the centroid of its own cluster, leave it in that cluster; otherwise, reassign it to the cluster whose centroid is closest to it.
- iv. Repeat step (iii) for each case.
- v. Repeat steps (ii), (iii), and (iv) until no cases are reassigned.

Dendrogram

Dendrogram, also called hierarchical tree diagram or plot, shows the relative size of the proximity coefficients at which cases are combined. The bigger the distance coefficient or the smaller the similarity coefficient, the more clustering involved combining unlike entities, which may be undesirable. Trees are usually depicted horizontally, not vertically, with each row representing a case on the Y axis, while the X axis is a rescaled version of the proximity coefficients. Cases with low distance/ high similarity are close together. Cases showing low distance are close, with a line linking them a short distance

Quantitative Methods for Social Sciences

from the left of the Dendrogram, indicating that they are agglomerated into a distance coefficient, indicating alikeness. When, the linking line is a bigh distance coefficient into a distance coefficient. a low distance coefficient, management at a high distance coefficient, indicating line is to the the Dendrogram the linkage occurs at a high distance coefficient, indicating to the anglomerated even though much less alike. If a similarity management of the V the Dendrogram the Inikage occurs clusters are agglomerated even though much less alike. If a similarity measure distance measure, the rescaling of the X axis still produces at the clusters are agglomerated even and gradient of the X axis still produces a diagonal state of the rather than a distance mean and low alikeness to the left and low alikeness to the right

Distance measures

Given two objects X and Y in a 'p' dimensional space, a dissimilarity measure and the space of t

 $d(X,Y) \ge 0$ for all objects X and Y 1.

2.
$$d(X,Y) = 0 X = Y$$

d(X,Y) = d(Y,X)3.

Condition (3) implies that the measure is symmetric so that the dissimilarity me that compares X and Y is same as the comparison for object Y verses X. Condition requires the measures to be zero, when ever object X equals to object Y. The object Y. are identical if d(X, Y) = 0. Finally, Condition (1) implies that the measure is it negative.

Some dissimilarity measures are as follows.

Euclidian distance

This is probably the most commonly chosen type of distance. It is simply the geometry distance in the multidimensional space. It is computed as,

$$d(X,Y) = \left\{ \sum_{i=1}^{p} (X_i - Y_i)^2 \right\}^{\frac{1}{2}} \text{ or }$$

In matrix form

d (X,Y)= d (X,Y)=
$$\sqrt{(X-Y)'(X-Y)}$$

X = (X1, X2, ..., Xp)Where

$$Y = (Y1, Y2, ..., Yp)$$

The statistical distance between the same two observations is of the form

d (X,Y) =
$$\sqrt{(X-Y)'A(X-Y)}$$
,

where A = S-1 and S contains the sample variances and covariances.

Euclidian and square Euclidian distances are usually computed from raw data and from standardized data

Square euclidean distance

Square the standard Euclidean distance in order to place progressively greater weight on objects that are further apart. This distance is computed as:

$$d^{2}(X,Y) = \left\{ \sum_{i=1}^{p} \left| X_{i} - Y_{i} \right|^{m} \right\}^{\frac{1}{m}}$$

or in matrix form

$$d^{2}(X,Y) = (X - Y)'(X - Y)$$

Minkowski metric

When there is no idea about prior knowledge of the distance group then one goes for Minkowski metric. This can be computed as given below:

$$d(X,Y) = \{\sum_{i=1}^{p} |X_i - Y_i|^m\}^{\frac{1}{m}}$$

For m = 1, d(X,Y) measures the city block distance between two points in p dimensions.

For m = 2, d(X,Y) becomes the Euclidean distance. In general, varying m changes the weight given to larger and smaller differences.

City-block (Manhattan) distance

This distance is simply the average difference across dimensions. In most cases, this distance measure yields result similar to the simple Euclidean distance. This can be computed as :

$$d(X,Y) = \sum_{i=1}^{p} |X_i - Y_i|$$

Chebychev distance

This distance measure may be appropriate in case when we want to define the objects as different if they are different on any one of the dimensions. The Chebychev distance is computed as:

$$d(X,Y) = maximum |X_i - Y_i|$$

Two additional popular measures of distance or dissimilarity are given by the Canberra metric and the Czekanowski coefficient. Both of these measures are defined for non negative variables only. We have

Canberra metric:
$$d(X, Y) = \sum_{i=1}^{p} \frac{|X_i - Y_i|}{(X_i + Y_i)}$$

Czekanowski coefficient = 1-
$$\frac{2\sum_{i=1}^{p} \min(X_i, Y_i)}{\sum_{i=1}^{p} (X_i - Y_i)}$$

ILLUSTRATION

ILLUST KALLON Given below is food nutrient data on calories, protein, fat, calcium and iron. The Given below is food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study is to identify suitable clusters of food nutrient data based on the study of the Given below is food nutrient data build be clusters of food nutrient data based on the objective of the study is to identify suitable clusters of food nutrient data based on the objective of the study and Collins, 1990). five variables (Chatfield and Collins, 1990).

| alories 340 | | | Calcium | 1 |
|----------------|----|----------------------|--------------|---|
| 210 | 20 | 28 | 9 | Iron |
| 245 | 21 | 17 | 9 | 2.6 |
| 420 | 15 | 39 | 7 | 2.7 |
| 375 | 19 | 32 | 9 | 2 |
| 180 | 22 | 10 | 17 | 2.6 |
| 115 | 20 | 3 | 8 | 3.7 |
| 170 | 25 | 7 | 12 | 1.4 |
| 60 | 26 | 5 | 14 | 1.5 |
| | 20 | 20 | 9 | 5.9 |
| 00 | 18 | 25 | 9 | 2.6 |
| | 20 | 23 | 9 | 2.3 |
| 40 | 19 | 20 | 9 | 2.5 |
| 40 | | 30 | 9 | 2.5 |
| 55 | 19 | | | 2.4 |
|)5 | 18 | 14 | 7 | 2.5 |
| 5 | 23 | 9 | 9 | 2.7 |
| 5 | 22 | 4 | 25 | 0.6 |
| | 11 | 1 | 82 | 6 |
| | 7 | 1 | 74 | 5.4 |
| | 14 | 2 | 38 | 0.8 |
| | 16 | 5 | 15 | 0.5 |
| | 19 | 13 | 5 | 1 |
| | 16 | 9 | 157 | 1.8 |
| | 16 | 11 | 14 | 1.3 |
| | | | | 0.7 |
| | | | | 2.5 |
| | | | | 1.2 |
| | | | | 2.6 |
| | | 17 22 25 23 | 22 9 25 7 | 22 9 367 25 7 7 |

| Table 1: Food nutrient | data on calories, pr | otein, fat, cal | cium and iron |
|------------------------|----------------------|-----------------|---------------|
| Table 1: F000 Huttrent | | | |

Analysis using SPSS

Start by entering the datasheet into SPSS using the steps below. Step: Go to file \rightarrow open \rightarrow browse the datasheet \rightarrow click open or Enter all the data in the data editor as shown in Figure 1.

Cluster Analysis

| le Edit | View Data | ransform | - | | | | | Unites | | | Vindow | - | | | | 1 | 1 | - | 下生亡 | The second | 1 | | 1.200 | | |
|-----------|-----------------------------------|------------|---------|---------|---|----------------|-------------------|---------------|-------|-----------|--------|-----------|----------|--|---|----------|----------------|--------|----------------|------------|-----------|-----------|----------------|------------|---------------|
| | | Kr. | 2 | | | | h | * | × | | 50 | | 14 | 6 | • | 485 | | | | | | | | | |
| - | | | | | and the second se | of the wide of | | and the state | | - martine | | | 1 Martin | and the second sec | | | | | | | | | | Visible: 6 | of 6 Variable |
| | Fooditems | Calor | ies | Protein | Fat | Cal | cium | iron | var | 1 | Var | | var | Var | | var | Var | 1 | Var | Var | | var | var | Va | |
| 1 | 1 | | 340 | 2 | 0 28 | | 9 | 2.6 | | | | | | | | | | | | | | | | | |
| 2 | 2 | | 245 | 2 | 1 17 | | 9 | 27 | | | | | | | | | | | | | | | | | |
| 3 | 3 | | 420 | 1 | 5 39 | | 7 | 2.0 | | | | | | | | | | | | | | | | | |
| 4 | 4 | | 375 | 1 | 9 32 | | 9 | 2.6 | | | | | | | | | | | | | | | | | |
| 5 | 5 | | 180 | 2 | 2 10 | | 17 | 3.7 | | | | | | | | | | | | | | | | | |
| 6 | 6 | | 115 | 2 | 0 3 | | 8 | 1.4 | | | | | | | | | | | | | | | | | |
| 7 | 7 | | 170 | 2 | | | 12 | 15 | | | | | | | | | | | | | | | | | |
| 8 | 8 | | 160 | 2 | 6 5 | | - 14 | 5.9 | | | | | | | | | | | | | | | | | |
| 9 | 9 | | 265 | 2 | 0 20 | | 9 | 2.6 | | | | | | | | | | | | | | | | | |
| 10 | 10 | | 300 | 1 | 8 25 | | 9 | 2.3 | | | | | | | | | | | | | | | | | |
| 11 | 11 | | 340 | 2 | 0 26 | | 9 | 2.5 | | | | | | | | | | | | | | | | | |
| 12 | 12 | | 340 | 1 | 9 25 | | 9 | 25 | | | | | | | | | | | | | | | | | |
| 13 | 13 | | 355 | 1 | 9 30 | | 9 | 2.4 | | | | | | | | | | | | | | | | | |
| 14 | 14 | | 205 | 1 | 8 14 | | 7 | 2.5 | | | | | | | | | | | | | | | | | - + |
| 15 | 15 | | 185 | 2 | 3 9 |) | 9 | 2.7 | | | | | | | | | | | | | | | | | |
| 16 | 16 | | 135 | 2 | 2 4 | | 25 | .6 | | | | | | | | | | 1 | | | | | | | |
| 17 | 17 | | 70 | | 11 1 | | 82 | 60 | | - | | | | | | | | | | | | | | | |
| 18 | 18 | | 45 | | 7 . | 1 | 74 | 5.4 | | | | | | | | | | | | | | | | | |
| 19 | 19 | | 90 | 1 | 14 3 | 2 | 38 | .8 | | | | | | | | | | | | | | | | | |
| 20 | 20 | | 135 | | 16 5 | 5 | 15 | 5 | | | | | | | | | | | | | | | | | |
| 21 | 21 | | 200 | | 19 13 | 3 | 5 | 1.0 | | | | | | | | | | | | | | | | | |
| 22 | 22 | | 155 | | 16 | 9 | 157 | 1.8 | | | | | | | | | 1 | | | | | | | | |
| 23 | 23 | | 195 | | 16 1 | 1 | 14 | 13 | | | | | | | | | | | | | | | | | |
| | 0 | the states | APR COM | - | and and | 100 | No. of Concession | | A low | - A.C. 14 | - | 2 million | - | action of | | States | Selection of | in the | and the second | and the | | 91 | - North Street | 1.000 | A - A - A |
| Deta View | Variable View | | | | | | | | | | | | | | | | | | | | | | | | |
| | Contraction of the local distance | | | | - | - | - | | - | | | 242 | | | 1 | a harris | and the second | 5.00 | | | | | | | TT |
| | | - | 100 | | 10.111 | _ | | 1.0000 | _ | - | - | | - | | - | | | | 1 | BM SPSS | Statistic | S PTOCESS | or is ready | | 11-04 AM |

Fig 1: Screen shot after entering the data in data editor

Now click Analyze \rightarrow Classify \rightarrow Hierarchical Cluster as shown in Figure 2.

| He Edit View | | Analyze Direct Marketing Graph Reports | | |
|---------------------|---|---|---|--------------------------|
| | | Descriptive Statistics | x x x x x x x x x x x x x x x x x x x | |
| | | Tables + | | Visible: 6 of 6 Variable |
| Fe | oditems Calori | Compare Means 🕨 🕨 | Inon var var var var var var var | var var |
| 1 1 | | General Linear Model 🔹 🕨 | 26 | |
| 2 2 | | Generalized Linear Models 🕨 | 2.7 | |
| 3 3 | | Mixed Models > | 2.0 | |
| 4 4 | | Correlate + | 26 | |
| 5 5 | | Regression + | 37 | |
| 6 6 | | Loginear > | 1.4 | |
| 1 1 | | Neural Networks | 1.5 | |
| 8 B 9 9 | | Classify + | TwoStep Cluster_ | |
| 9 9 10 10 | | Dimension Reduction | K-Heans Cluster | |
| 11 11 | | Scale + | Herarchical Cluster | |
| 12 12 | | Nonparametric Tests | | |
| 13 13 | | Forecasting + | | |
| 14 14 | | Survival > | Discriminant | |
| 15 15 | | Multiple Response | Mearest Neighbor | |
| 16 16 | | Missing Value Analysis | 6 | |
| 17 17 | | Multiple Imputation + | 6.0 | |
| 18 18 | | Complex Samples | 54 | |
| 19 19 | | Quality Control | 8 | |
| 20 20 | | ROC Curve | 5 | |
| 21 21 | | 200 19 13 | 5 1.0 | |
| 22 22 | | 155 16 9 1 | 7 18 | |
| 23 23 | | 195 16 11 | 4 13 | |
| 4 | and the state of the | Salar Branks Schuler State | | Constant of the Constant |
| Data View Vata | dela Viene | | | |
| | N M LONG | and a company planter in | a company with the second of the second of the second | Consider and and |
| erarchical Clust | H | | IBM SPSS Statistics Proces | seor is ready |

Fig 2: Screen shot of selecting the analysis procedure

Then Identify Name as the variable by which to label cases and Calories, Protein, Fat, Calcium, and Iron as the variables. Indicate that you want to cluster cases rather than variables and want to display both statistics and plots as shown in Fig 3.

| | Variables(s): |
|--|--|
| Hierarchical Cluster Analysis | Calories Protein Fat Calcium Iron |
| | Label Cases by: Foodterns Cluster Cases O Variables |
| OK Raute | Display Statistics Plots Reset |
| r - 3. Cluster cases rather than varia | bles and want to display both statistics and plots |

Click Statistics and indicate that you want to see an Agglomeration schedule with?

| Hierarchical Ouster Analysis: Statistics | Dendrogram |
|---|--|
| Aggiomeration schedule Paraimity matrix Cluster Membership O None | Icicle |
| Single solution Number of clusters Range of solutions | By: 1 |
| Minimum number of clusters: 2 Magimum number of clusters: 6 | Orientation |
| | Continues Carcel Hat |

Fig 4: Hierarchical cluster analysis statistics

Fig 5: Hierarchical cluster analysis plot

5 2

4, and 5 cluster solutions. Click Continue as shown in Fig 4

Click plots and indicate that you want a Dendogram and a verticle Icicle plot with and 4 cluster solutions. Gut a second and 4 cluster solutions. Click Continue as shown in Fig 5 Click Method and indicate that you want to use the Between-groups linkage method clustering, squared Fuelidian ti clustering, squared Euclidian distances, and variables standardized to z scores (

variable contributes equally). Click Continue as shown in Fig 6.

Cluster Analysis

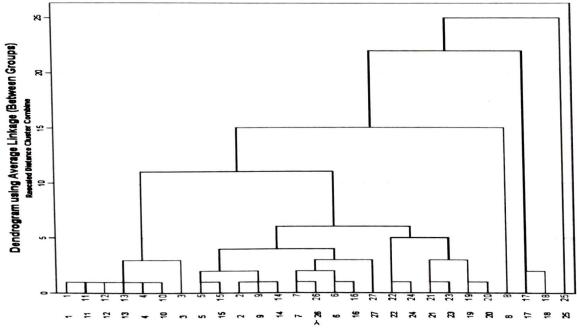
| Cluster <u>M</u> ethod Measure | Between-groups linkage | Contraction of the | Cluster Membership | | | | | |
|-----------------------------------|---------------------------|----------------------|-------------------------------|--|--|--|--|--|
| linterval: | Squared Euclidean distant | C0 T | O Single solution | | | | | |
| | Power 2 * Root | 2 - | Number of clusters | | | | | |
| O Counts | Chi-squared measure | a | Range of solutions | | | | | |
| O Binary | Squared Euclidean distan | C6 • | Minimum number of clusters: 2 | | | | | |
| | Present 1 5bse | nt o | Maximum number of clusters: 6 | | | | | |
| Transform Va | lues | Transform Measure | | | | | | |
| Standardize | Z scores | Absolute values | | | | | | |
| | By variable | Change sign | | | | | | |
| | O By case: | Rescale to 0-1 range | | | | | | |

Fig 6: Hierarchical cluster analysis method

Click Save and indicate that you want to save, for each case, the cluster to which the case is assigned for 2, 3, 4, 5 and 6 cluster solutions. Click Continue, OK as shown in Fig 7

SPSS starts by standardizing all of the variables to mean 0, variance 1. This results in all the variables being on the same scale and being equally weighted.

Dendrogram



Interpretation

The main objective of our analysis is to group the food items on the basis of their nutrient content based on the five variables such that food items with in the groups are homogeneous and between the groups are heterogeneous.

| Table 2: Interpret | ation |
|--------------------|---|
| Number of groups | Food items |
| Two groups | Group-1 (1,11,12,,18) Group-2 (25) |
| Three groups | Group-1 (1,11,,8) Group-2 (17,18) Group-3 (25) |
| Four groups | Group-1 (1,11,,20) Group-2 (8) Group-3 (17,18) Group-4 (25) |
| Five groups | Group-1 (1,11,,3) Group-2 (5,15,,20) Group-3 (8) Group-4 (17,18) Group-5 (25) |
| Six groups | Group-1 (1,11,,3) Group-2 (5,15,,27) Group-3 (22,24,20) Group-4 (8) Group-5 (17,18) Group-6 (25) |

Illustration (Using survey data from social science)

Given below is a part of the data based on a study which was conducted to understand the socio-economic implication of climate and vulnerability of farmers in and ecosystem of Rajasthan by Sarkar (2014). Two districts Jodhpur and Jaisalmer wer selected from arid ecosystem and 100 farmers were selected randomly for the present study. However, for the present chapter, in order to demonstrate the similarity in terms of adaptive behaviour of the farmers, the cluster analysis was performed by considering variables like awareness, attitude towards climate change, egalitarianism, risk perception w.r.t. 20 farmers.

Table 3: Illustration

| | | | | - tial |
|-------------|-----------|----------|-------------|-----------------------|
| Farmers' ID | Awareness | Attitude | Egalitarism | Risk perception 60 |
| 1 | 26 | 60 | 37 | 58 |
| 2 | 18 | 43 | 25 | 65 |
| 3 | 25 | 67 | 40 | 57 |
| 4 | 23 | 53 | 34 | 41 |
| 5 | 20 | 41 | 37 | 50 |
| 6 | 16 | 37 | 38 | 65 |
| 7 | 23 | 60 | 38 | |

| Farmers' ID | Awareness | Attitude | Egalitarism | Risk perception |
|-------------|-----------|----------|-------------|------------------------|
| 8 | 19 | 41 | 27 | 50 |
| 9 | 23 | 41 | 26 | 64 |
| 10 | 26 | 61 | 37 | 60 |
| 11 | 18 | 48 | 25 | 59 |
| 12 | 26 | 67 | 40 | 67 |
| 13 | 23 | 53 | 35 | 57 |
| 14 | 20 | 41 | 37 | 41 |
| 15 | 16 | 37 | 38 | 48 |
| 16 | 25 | 59 | 38 | 66 |
| 17 | 19 | 40 | 27 | 50 |
| 18 | 23 | 42 | 27 | 64 |
| 19 | 26 | 68 | 36 | 61 |
| 20 | 16 | 42 | 25 | 58 |

Cluster Analysis

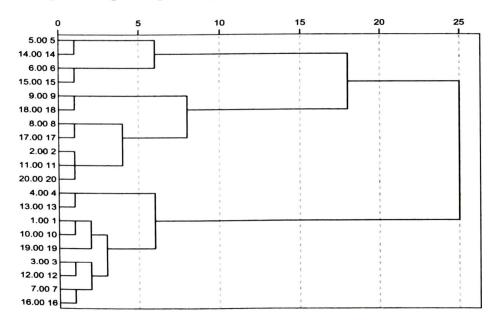
Here, the purpose of the cluster analysis is to group the farmer based on their adaptive behaviour so that appropriate action can be suggested for the farmers who are lagging behind. Two groups were formed viz. adaptors and non-adaptors. The results are summarized as follows:

Table 4: Adopters and non-adopters

| Groups | Farmers' ID |
|--------------|-----------------------------|
| Adopters | 1,3,4,7,10,12,13,16 and 19 |
| Non-adopters | 2,5,6,8,9,11,14,15,17,18,20 |

Dendogram

Dendrogram using Average Linkage (Between Groups) Rescaled Distance Cluster Combine



REFERENCES

5

- Afifi, A., V. A. Clark and S. Marg (2004), Computer Aided Multivariate Analysis. CRC Press, USA.
- Chatfield, C. and A. J. Collins (1990), Introduction to Multivariate Analysis. Chapman and Hall Publications.
- Johnson, R. A. and D. W. Wichern (2006), Applied Multivariate Statistical Analysis. 5th Edn., London, Inc. Pearson Prentice Hall.
- Sarkar, S. (2014), Assessment of Climate Change Led Vulnerability and Simulating the Adaptive Behaviour of Farmers in the Himalayan and Arid Ecosystems. Ph.D. Thesis, IARI, New Delhi.