Segmenting the milk production in the state of Tamil Nadu (India) into homogenous milk zones: A Multi Dimensional Scaling Approach

A.Serma Saravana Pandian¹, K.N.Selvakumar² and M.Prabu³
Tamil Nadu Veterinary and Animal Sciences University, Chennai- 600 007, India.
pandian23@gmail.com; knselva@yahoo.com; mprabu23@gmail.com

Abstract: In this study an attempt was made to segment milk production in the state of Tamil Nadu into homogenous milk zones based on the resource endowments favourable for the development of dairy industry. Based on the results of multi dimensional scaling, four distinct groups of districts were carved out based on their proximity and the districts were identified in different zones. In the first zone, there were seven districts as Kanchipuram, Vellore, Thiruvannamalai, Villuppuram, Pudukottai, Namakkal and Salem districts. In the second zone there were five districts namely Erode, Coimbatore, Dharmapuri, Didigul and Thiruchirappalli districts. The third zone comprises four districts viz., Thoothukudi, Tirunelveli, Virudhunagar and Nilgiri districts. The districts in the Zone IV included Nagappattinam, Thanjavur, Cuddalore, Madurai, Sivagangai, Ramanathapuram and Kaniyakumari districts. The multiple discriminant function analysis was used to check whether the groups of districts (Zones) are homogenous or not. The results of the Eigen values and percentage variance explained by the three discriminant functions showed that the first function has the highest percentage of variance (77.92 per cent), followed by the second function (21.47 per cent) and the third function has the least percentage of variance (0.607 per cent). These three functions together could explain 100 per cent of the variation and the estimated discriminant functions have effectively discriminated the districts in four zones and the districts in each of the four zones were of homogenous nature.

Keyword: Milk zones, segmenting, Multi Dimensional Scaling.

Introduction

Dairying in India has emerged as an important sub-sector, accounting for nearly two-thirds of the total livestock contribution to GNP with an encouraging growth rate of five per cent over years. Through improved breeding, feeding and management programmes, there has been marked improvement in the country’s milk production and productivity of milch animals. Tamil Nadu, the southernmost state of India is one of the top ten milk producing states in the country with an annual milk production of 4.75 million ones (Statistical Handbook of Tamil Nadu, 2004). Dairy farming as visualized by the farmers in Tamil Nadu state is it is part of an integrated agricultural system where dairy and agriculture complement each other. To have an effective planning and policy making for the development of dairy industry in the state, regional planning or zone based planning is important. For this purpose, segmenting should be carried out based on the resource endowment favourable for the dairy industry. Further, the identified zones should be homogenous. Keeping these facts in mind the present study was designed to segment milk production in the state of Tamil Nadu into homogenous milk zones.

Data and methodology

Data

The data for the study were collected from the secondary sources like the reports of Directorate of Economics and Statistics and Directorate of Veterinary Services, Govt. of Tamil Nadu, Chennai.

Analytical Tools Applied

By applying multidimensional scaling method, the state of Tamil Nadu was segmented into four homogenous milk zones based on the resource endowment of the districts favorable for dairy development. Then multiple discriminant function analysis was used to check whether the groups of districts (Zones) are homogenous or not.

Multidimensional Scaling

Multidimensional scaling (MDS) is an important analytical tool for the purpose of grouping. The goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities among the investigated objects (Karamathullah et al.,2002).

MDS is a procedure to "rearrange" objects in an efficient manner, so as to arrive at a configuration that best approximates the observed distances. It actually moves objects around in the space defined by the requested number of dimensions, and checks how well the distances between objects can be reproduced by the new configuration. In more technical terms, it uses a function minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit or minimizing "lack of fit". The most common measure that is used to evaluate how well a particular configuration reproduces the observed distance matrix is the stress measure. The raw stress value Phi of a configuration is...
### List of Variables taken for analysis

<table>
<thead>
<tr>
<th>Sl. NO</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Milk Production per 100 sq.km of the district (in 000 tonnes)</td>
</tr>
<tr>
<td>2</td>
<td>Indigenous cattle population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>3</td>
<td>Cross bred and Exotic cattle population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>4</td>
<td>Native purebred cattle population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>5</td>
<td>Indigenous Buffalo population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>6</td>
<td>Graded Buffalo population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>7</td>
<td>She-Buffalo population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>8</td>
<td>Veterinary Institution per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>9</td>
<td>Financial Institution per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>10</td>
<td>Agricultural labour population per 100 sq.km (in numbers)</td>
</tr>
<tr>
<td>11</td>
<td>Area under permanent pasture per 100 sq.km (in ha)</td>
</tr>
<tr>
<td>12</td>
<td>Area under Paddy per 100 sq.km (in ha)</td>
</tr>
<tr>
<td>13</td>
<td>Area under Maize per 100 sq.km (in ha)</td>
</tr>
<tr>
<td>14</td>
<td>Area under Groundnut per 100 sq.km (in ha)</td>
</tr>
<tr>
<td>15</td>
<td>Area under other cereals per 100 sq.km (in ha)</td>
</tr>
<tr>
<td>16</td>
<td>Gross cropped area per 100 sq.km (in ha)</td>
</tr>
<tr>
<td>17</td>
<td>Average Annual Rainfall (in mm)</td>
</tr>
</tbody>
</table>

**Fig. 1.** Multi Dimensional Scaling Configuration of districts in the context of milk production - 3 D

In this formula, \( d_{ij} \) stands for the reproduced distances, given the respective number of dimensions, and \( \partial_{ij} \) (delta_ij) stands for the input data. The expression \( f (\partial_{ij}) \) indicates a non-metric, monotone transformation of the observed input data. Thus, it will attempt to reproduce the general rank ordering of distances between the objects in the analysis. Smaller the stress value, the better is the fit of the reproduced distance matrix to the observed distance matrix. The interpretation of dimensions usually represents the final step of the analysis. As mentioned earlier, the actual orientations of the axes from the MDS analysis are arbitrary, and can be rotated in any direction. Three-dimensional solutions can be illustrated graphically.

**Discriminant function analysis**

In this study the discriminant function analysis was used to determine which variables discriminate between two or more naturally occurring groups and to check whether the groups are homogenous or not. (Karamathullah et al., 2002). When there are more than two groups, then we can estimate more than one discriminant function, which could be called as discriminant functions for multiple groups.

The result has given 'b' regression coefficients for each variable in each discriminant function, and they can be interpreted as; larger the standardized coefficient, the greater is the contribution of the respective variable to the discrimination between groups.

Discriminant function analysis is computationally very similar to MANOVA (Multivariate analysis of variance). It was assumed that the data (for the variables) represent a sample from a multivariate normal distribution. It is assumed that the variance/covariance matrices of variables are homogeneous across groups. Another assumption of discriminant function analysis is that the variables that are used to discriminate between groups are not completely redundant.

The data for the selected variables were tabulated and the discriminant function analysis was run using, SPSS 9.0R for Windows R, a computer package for the statistical analysis in social sciences. The linear discriminant function used for the study is of the following form.

\[
Z = b_1 X_1 + b_2 X_2 + \ldots + b_i X_i \quad (i = 1, 2, 3\ldots17)
\]

where

- \( Z \) = Total discriminant score
- \( X_i \) = Variables
- The variables refer to the 17 numbers of dairy related variables which were used in Multi dimensional scaling. \( (X_1 \ldots X_{17}) \) (Milk production \ldots Average annual rainfall).

**Results and discussion**

**Multidimensional scaling**

Based on the dairy related variables for each districts of Tamilnadu, multi dimensional scaling...
was carried out. The result of the multi-
dimensional scaling is depicted in Fig.1. which
examines configuration of districts based on three
dimensions. This helps to identify the similar
districts. Multi-dimensional scaling moves objects
around the space defined by the requisite number
of dimensions and checks how well the distances
between the objects can be reproduced by new
configuration. The goodness of fit is tested through
the statistic called stress and the Stress
value is 0.30005.

Table 1. Milk Zones identified through
Multidimensional Scaling

<table>
<thead>
<tr>
<th>SL NO</th>
<th>Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone I</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Kanchipuram</td>
</tr>
<tr>
<td>2</td>
<td>Vellore</td>
</tr>
<tr>
<td>3</td>
<td>Thiruvannamalai</td>
</tr>
<tr>
<td>4</td>
<td>Villuppuram</td>
</tr>
<tr>
<td>5</td>
<td>Pudukkottai</td>
</tr>
<tr>
<td>6</td>
<td>Namakkal</td>
</tr>
<tr>
<td>7</td>
<td>Salem</td>
</tr>
<tr>
<td>Zone II</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Erode</td>
</tr>
<tr>
<td>2</td>
<td>Coimbatore</td>
</tr>
<tr>
<td>3</td>
<td>Dharmapuri</td>
</tr>
<tr>
<td>4</td>
<td>Dindigul</td>
</tr>
<tr>
<td>5</td>
<td>Thiruchirapalli</td>
</tr>
<tr>
<td>Zone III</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Thoothukudi</td>
</tr>
<tr>
<td>2</td>
<td>Tirunelveli</td>
</tr>
<tr>
<td>3</td>
<td>Virudhunagar</td>
</tr>
<tr>
<td>4</td>
<td>Nilagiri</td>
</tr>
<tr>
<td>Zone IV</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Nagappattinam</td>
</tr>
<tr>
<td>2</td>
<td>Thanjavur</td>
</tr>
<tr>
<td>3</td>
<td>Cuddalore</td>
</tr>
<tr>
<td>4</td>
<td>Madurai</td>
</tr>
<tr>
<td>5</td>
<td>Sivagangai</td>
</tr>
<tr>
<td>6</td>
<td>Ramanathapuram</td>
</tr>
<tr>
<td>7</td>
<td>Kanniyakumari</td>
</tr>
</tbody>
</table>

On analyzing the orientation of the districts in
three dimensional view of the Fig.1, four distinct
groups of districts were carved out based on their
proximity or closeness and the districts identified in
different zones are given in Table 1. From the
Table it could be observed that there are four
distinct groups of districts. In the first zone there
were seven districts as Kanchipuram, Vellore,
Thiruvannamalai, Villuppuram, Pudukkottai,
Namakkal and Salem which comprises mainly the
north-eastern and north western agro-climatic
zone of Tamilnadu except pudukkottai district. In
the second zone there were five districts namely
Erode and Coimbatore districts which cover the
western agro climatic zone, Dharmapuri, Dindigul
and Thiruchirappalli districts which are the
components of different agro climatic zones of the
state. The third zone comprises four districts viz.,
Thoothukudi, Tirunelveli and Virudhunagar which
are the component districts of southern agro
climatic zone and Nilagiri district which is a hilly
agro climatic district. The districts in the Zone IV
included Nagappattinam, Thanjavur, Cuddalore,
Madurai, Sivagangai, Ramanathapuram and
Kanniyakumari districts which represent Cauvery
delta, southern and high rainfall agro-climatic
zones of Tamilnadu.

Discriminant Function Analysis

The multiple discriminant function analysis was
used to check whether the groups of districts
(Zones) are homogenous or not. (Karamathullah et
al.,2002). For this purpose the seventeen numbers
of dairy related variables that were used for zoning
were separately taken for each district under the
four different zones. The data were segregated
and taken under four groups and subjected to
discriminant function analysis. The canonical
correlation analysis would determine the
successive functions and canonical roots. The
results of the discriminant function analysis are
given in Table 2, 3 and 4.

Table 2 provides the co-efficients of discriminant
functions for each dairy related variable with
respect to the three estimated discriminant
functions. From the values of the co-efficients in
the first function it could be inferred that the
variables such as area under groundnut, native
purebred cattle population, indigenous population
and the total she buffalo population have the
highest contribution and hence these variables in
the first function can effectively discriminate
between the districts of the four zones. In the
second function, the variables number of
veterinary institutions, Indigenous cattle population
and native purebred population have better
chance to discriminate the zones. While in the
third function, the variables area under maize has
the highest co-efficient value followed by the
variables, average annual rainfall, area under
permanent pasture, the area under other cereals
and the agricultural labour population.

The results of the Eigen values and percentage
variance explained by the three discriminant
functions are given in Table 3. On observing this
By using multi dimensional scaling, 4 distinct groups of districts were carved out based on their proximity. The multiple discriminant function analysis was used to check whether the groups of districts (Zones) are homogeneous or not. The results of the Eigen values and % variance explained by the three discriminant functions showed that the first function has the highest % variance (77.92 %), followed by the second function (21.47 %) and the third function has the least percentage of variance (0.607 %). These three functions together could explain 100 % of the variation.

Based on the value obtained from the discriminant functions for each district, they were classified to check the ability of the discriminant functions in effectively discriminating the zones. The result is given in Table 4 and the % of cases correctly classified by the model is also given in the table. On perusal of the table it could be inferred that the estimated discriminant functions have effectively discriminated the districts in 4 zones and the districts in each of the 4 zones are of homogenous nature.

Summary and conclusion

By using multi dimensional scaling, 4 distinct groups of districts were carved out based on their proximity. The multiple discriminant function analysis was used to check whether the groups of districts (Zones) are homogeneous or not. The results of the Eigen values and % variance explained by the three discriminant functions showed that the first function has the highest % variance (77.92 %), followed by the second function (21.47 %) and the third function has the least percentage of variance (0.607 %). These three functions together could explain 100 % of the variation and the estimated discriminant functions have effectively discriminated the districts in 4 zones and the districts in each of the 4 zones were of homogenous nature.

References