Tropical Agricultural Research Vol. 13: 156-164 (2001)

Evaluation of Performance of Public Schools: An Application of GIS in Education Surveys

A.J. Satharasinghe, R.O. Thattil[1], S. Samita[1] and R.P. De Silva[2]

Postgraduate Institute of Agriculture
University of Peradeniya
Peradeniya, Sri Lanka

ABSTRACT. Stratification is a widely used sample survey technique. In stratified sampling, the sampling frame is divided into strata so that the study variables vary less within strata than in the un-stratified population. Independent samples are then drawn from the strata. District boundaries and type of school may not be the right variables to stratify student population for the surveys designed to evaluate performance of students. This study is an attempt to design suitable criteria for stratifying the area and schools and thereby to introduce a new efficient sampling design to be used in evaluation of performance of schools and students.

A geographical buffering technique is introduced in a Geographic Information Systems (GIS) environment with selected spatial variables, in order to construct homogenized regions. Spatial variables such as access to facilities including towns, main roads, hospitals, geographic distribution of poverty, population density etc., are used to construct these regions. Non-spatial data of schools including availability of physical resources, characteristics of teachers and students are used to classify the schools. These two classifications provide strata for two stage stratified sampling designs. The spatial and non-spatial attribute variables are combined to classify schools to be used for one stage stratified sampling design. Geographic Information Systems is used to derive spatial variables. Classification is done by multivariate techniques and using the Extended Ekman rule with necessary adjustments. The utility of new classification for future research on education for Southern Province is explained.

The study is currently ongoing. The results of work done so far are presented here.

INTRODUCTION

The government has decided to restructure the education system on an urgent priority basis as an integral part of the development process of the country. In this effort, much emphasis has been placed on basic education. Education reforms by stages are being implemented with the objective of improving basic competencies of children. In order to evaluate the impact of reforms, the entire education system in our country needs to be reviewed, researched and monitored.

---

1 Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.

2 Department of Agricultural Engineering, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.

Estimates of various parameters that measure the performances of schools and students are required to evaluate the impact of these measures. Generally these parameters are estimated through sample surveys, as a census requires a large amount of financial and human resources. It is needless to emphasize the necessity to use a valid sampling design for these surveys.

The commonly used sampling design is stratified sampling design. In stratified sampling, we work with two sets of variables: study variables and stratification variables. It is true that the stratified sampling scheme provides the best estimates in that the associated standard errors are low. But the use of district and the type of school as the stratification variables in these surveys is questionable. The purpose of stratification is to group study units into homogeneous groups by using one or more variables related to the parameter(s) being estimated. Once the population is stratified, it is possible to draw a more representative sample and this contributes to lower the standard error of the estimates. However, in reality, the factors that are likely to influence the performance of schools within districts are diverse and vary widely. According to the stream of subjects taught and up to what grades of classes are available, schools of Sri Lanka are categorized into four types, although this classification is not strictly adhered to. Therefore, it is required to select more appropriate variable(s) to be used in stratification.

Many factors such as characteristics of pupils, parents, teachers and learning environment within the school, home, and community influence the performance of schools and students. Some of these characteristics can be described by spatial variables and others, by non-spatial attribute variables. In a review of available descriptive secondary data, Byrne (1959) noted the very high degree of correspondence between the social characteristics of areas and achievements of pupils in state secondary schools drawn from those areas. A characteristic of some spatial variables is that their values vary according to the distance from the sources.

Conventional statistical analysis either completely ignores spatial variation among observation units or assumes the existence of rigid regional boundaries within which the assumption of independence and randomness among observations holds. In reality, the degree of spatial distribution often changes continuously with the distance between observations. Spatial variables of units are not independent and closely located units can be more closely related than those farther apart. If this is not taken into account, models may end up being specified incorrectly, parameter estimates may be biassed and inferences may be incorrect. As such a method should be devised to incorporate effects of spatial variables in the process of stratifying the population. There can be many spatial variables influencing the performances of schools and students. This study uses two sets of spatial variables. One is distances to the facilities such as towns, Type A roads, hospitals *etc.* The other one is geographic distribution of poverty levels and population density. *GIS* is used to derive the spatial distribution of these variables. This study demonstrates an example, where GIS is used in educational research.

Spatial variables are used in this study to stratify the area. Non-spatial attribute data are used to stratify schools. As explained earlier, classification of schools by type cannot be justified for two reasons. Firstly the definition is not strictly adhered to. Secondly, type alone does not account either for facilities available in the schools or important characteristics of the school and children. Therefore this study uses a set of non-spatial attribute variables related to school performance to construct school strata. These strata can be used as first stage and second stage strata in a Two Stage Stratified Sampling Design.

The objective of this study is to propose a suitable sampling design for assessing the performances of schools and students of the Southern Province.

MATERIALS AND METHODS

Spatial variables

Digital maps listed below were used to derive spatial data.

1. Maps of Southern Province showing the geographic features such as boundaries of Divisional Secretary (DS) divisions, Type A, B, C and D roads, railway stations, hospitals, police stations and schools. Schools were mapped by creating points on the Southern Province boundary map, corresponding to the coordinates recorded on a *Global Positioning Systems (GPS)* unit.

2. Maps with data on poverty and population density by main cities/villages and DS divisions respectively of the Southern Province.

Hardware and software developments have made GIS a user-friendly environment. This study used MapInfo (ver. 4.0) and ArcView (ver. 3.2) GIS software, to derive the spatial variables.

Forest areas were removed from the land area map of the Southern Province. In order to subdivide the map of Southern Province into a set of homogeneous regions (strata) with respect to the variables influencing the performances, it is required to create a grid map of Southern Province consisting of small area units called cells (Fig. 1). Generally, square cells are used for this purpose (Solano *et al.*, 1995). The size of the cell has to be decided according to the average number of schools falling into a cell. If the average number of schools in a cell is large, the values of the spatial variables derived for the cell in which these schools are located will be the same for all schools in the cell. This will badly hamper the detection of spatial relationships and patterns. Average number of schools falling into a cell was computed for different cell sizes and the cell size corresponding to lowest number of schools was chosen as the cell size for the grid.

One of the spatial variables selected is the distance from cells to the nearest Type A road. The roads of the country are classified into four categories: Type A, B, C and D. Type A roads are wide tarred main roads. Using the GIS software functionality, distances for each and every point on the map to the nearest Type A road were computed and converted into a 10-point scale. The grid was overlaid on the map. The values of the distances calculated were transferred to the corresponding cells (Fig. 2).

Similarly distances to the other nearest facilities namely hospitals, police stations, towns, railway stations, forests (as a measure of remoteness) and distribution of poverty levels and population densities *etc.* were computed and transferred to the grid. In addition to this, average scholarship examination marks of the schools were also computed and transferred to the corresponding cells in the grid.

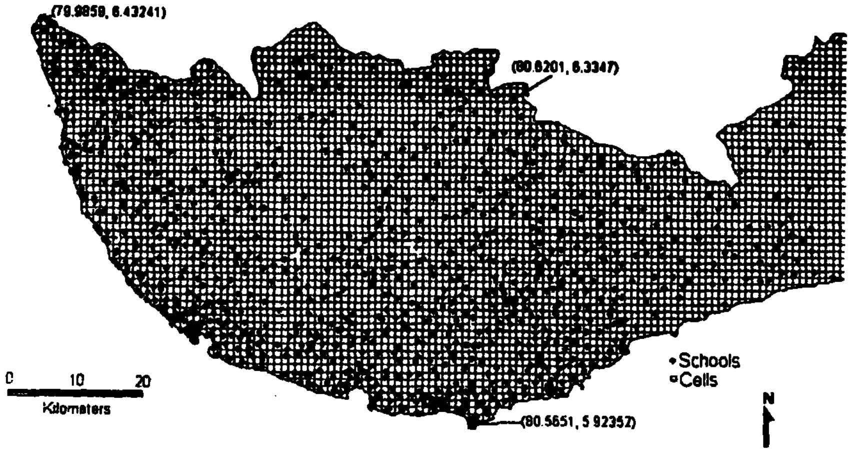

Fig. 1. A section of the 1 km grid and schools of the Southern Province.

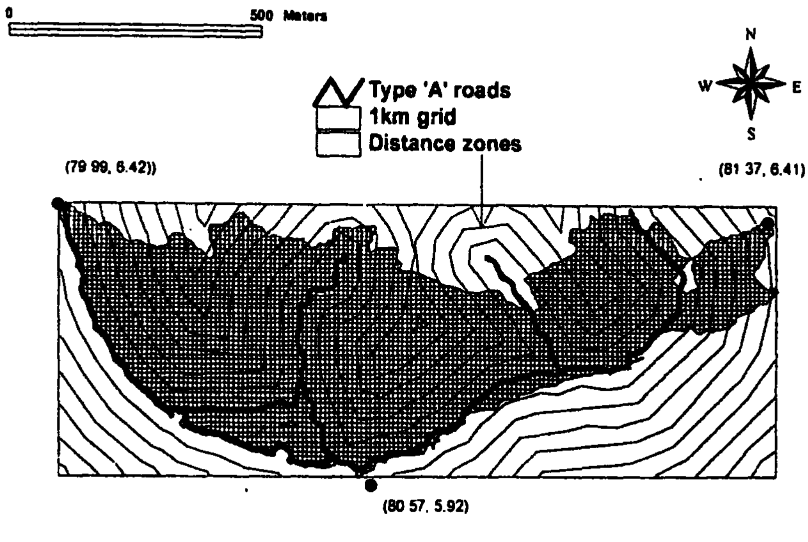

Fig. 2. Type 'A' roads, 1 km grid and distance zones.

The distances considered here are aerial distances. The Southern Province which is a flat area to a larger extent compared to other provinces, was chosen for this study as the aerial distances would give an indication on the extent to which the cells are closer to the facilities. The most appropriate way of doing this is to compute the distance along the network of roads by the network analysis (Green et al., 1998; Naude, 1995). Yet to handle such a large number of data points, the available hardware and GIS software are not powerful enough. As this study is an exploratory study aiming at introducing a GIS application in sampling and characteristics of the topology of the study area, which in this case is largely flat, aerial distances were used, as it is not likely to affect the findings significantly.

The values of spatial variables recorded in the cells of the grid were used to stratify the area of the Southern Province to form a set of homogenized areas in which the effect of the above variables on performance was the same or fell within a known range. Cells in the grid were then disassembled and converted into a database file together with location identities. In order to display the results of the analysis back on a map, location identities were also recorded in the data file.

Non-spatial attribute variables

The following non-spatial attribute variables were used to stratify schools. Data were obtained from the database of the annual school censuses of the Ministry of Education.

- Availability of libraries, computers, drinking water, electricity and sanitary facilities
- Total revenues received
- Percentages of teachers by qualification
- Percentage of teachers with professional qualifications
- Numbers of students of year five grade and of all grades
- Total numbers of classes of year 5 grade and of all grades
- Travel time from the school to the nearest urban centre (This is a spatial variable. Yet due to its importance at school level it was also used together with the non-spatial attribute variables)

Methods of stratification

Traditionally, stratification is done with respect to one or two variables. This does not guarantee that stratification has been done with respect to all important effects that could influence the study variable. Stratification by several variables lead to difficulties both administratively as well as in computations. This study employs two methods of multivariate stratification without affecting characteristics of stratified sampling.

Method 1: Multivariate stratification by multivariate procedures

Here, two multivariate procedures known as factor analysis and cluster analysis were used. After analysing the correlations, variables to be used for stratification were submitted to factor analysis. Factor analysis identifies underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction, by identifying a small number of factors, which explain most of the variances observed in a much larger number of variables (Unwin *et al.*, 1996). A sufficient number of factors were extracted and factor scores were computed.

Factor scores were submitted to cluster analysis. The cluster analysis identifies relatively homogeneous groups of cases based on factor scores. These groups were used as strata.

Applying these procedures for spatial variables and non-spatial attribute variables respectively the area of the Southern Province and schools were stratified.

Method 2: Multivariate stratification by using extended Ekman rule

A paper published by Dan (2000) extends the Ekman rule, which is a well-known method for univariate stratification. It approximately minimizes the variance of the estimates. This rule was applied for multivariate stratification as follows:

Correlation analysis and regression analysis were used to identify the factor that is most significantly related to the performances of the school. This factor was used, as the stratification variable and stratum boundaries were determined using the Extended Ekman Rule. An algorithm written in SAS software was used to determine boundaries.

RESULTS AND DISCUSSION

Determination of grid cell size

Number of schools falling into a cell decreases rapidly with the cell size. Grid size of 1 km × 1 km was chosen as in this grid, on an average, there is only one school in a cell (Fig. 3). Further reduction of cell size was not possible due to the limitations in computer hardware and software.

Stratification of the area

Correlation analysis revealed that average marks were significantly correlated with all spatial variables. Other than a few exceptions, spatial variables were significantly correlated with each other, at 95% level of confidence. Therefore, spatial variables were submitted to factor analysis. The variable of distance from the forest was excluded based on the communalities and factor loading statistics.

There were two factors with eigen values greater than one. These two factors were extracted using the Principal Component method together with Varimax rotation. These two factors accounted for 66% of the variation of the original five spatial variables. Factor 1 mainly represented the distance to the nearest police station; hospital and railway station while Factor 2 represented the distance to the nearest Type A road and to the town. Component plot in rotated space justifies the use of Principal Component Factor Extraction method and Varimax Rotation, as factor loadings are concentrated at the ends of zero lines.

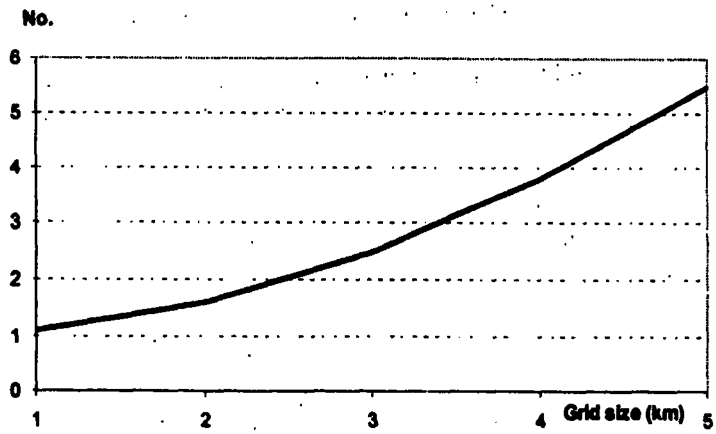

Fig. 3. Average number of schools per cell by cell size.

Factor scores were submitted to k means cluster analysis procedure. ANOVA tables of factor scores using clusters as groups and practical considerations supported the choice of 4 clusters. Therefore, cells were classified into four groups. Cells were reassembled with the cluster membership data and transferred to the grid. Cells were then coloured according to the cluster membership displaying the new homogeneous sub regions, which can be used as the first stage strata (Fig. 4). Boundaries of the new sub regions are very clear and are different from the administrative boundaries. One cluster consisted of a small number of cells, which are located very close to the forest area possibly indicating that it is the worst area with respect to facilities. The boundaries of the new regions reflect the characteristic of how far they are from the facilities that are likely to influence the performances of students..

Stratification of schools

Similar to the classification of area cells, factor and cluster analysis were run on non-spatial attribute data of schools. Different cluster sizes were compared and five clusters were selected as it described the data best, in terms of average and variability of marks within clusters. The number of schools in the clusters 1, 2, 3, 4 and 5 were 180, 411, 107, 149 and 140 respectively, whereas number of schools under Type 1AB, Type 1C, Type 1 and Type 2 were 80, 290, 457 and 339 respectively. These distributional differences were probably due to the fact that in grouping schools by type, only subjects taught and up to what grade classes are conducted were considered, whereas in this classification method, many important variables likely to influence the performances were considered. These clusters could be used as the second stage strata.

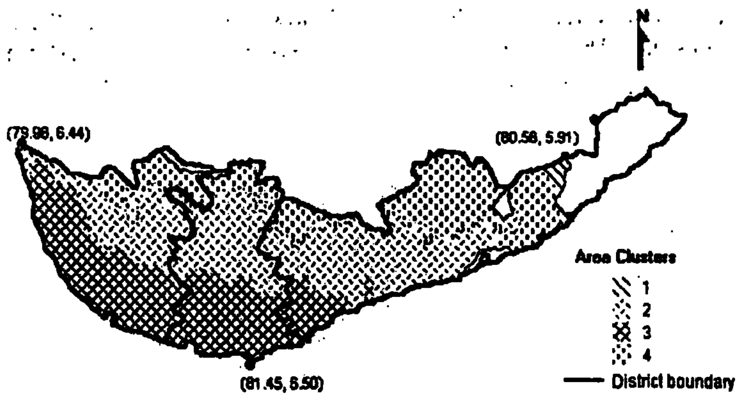

Fig. 4. Stratification of area by cluster membership.

Evaluation of sampling designs

The objective of this study is to stratify the area and the schools in a better way than using district and type of schools as the stratification parameters so that the new stratification system can be used for future research on education. This study tests the validity of the new system comparing estimates and associated standard errors derived from several sample designs.

Marks of year 5-scholarship examination of year 2000 are used as the performance variable. Samples are drawn according to the following sampling designs. Average performance and their standard errors are estimated. Standard errors are compared to propose the most suitable stratification system and sampling design.

a. Two-stage stratified sampling using the district and type of school as the first and second stage strata.
b. Two-stage stratified random sampling using the strata defined by Method 1.
c. Two-stage stratified random sampling using the strata defined by Method 2.
d. One-stage stratified random sampling by using strata defined by applying Method 1 to pooled spatial and non-spatial attribute data.
e. One-stage stratified random sampling by using strata defined by applying Method 2 to pooled spatial and non-spatial attribute data.

CONCLUSIONS

This study shows that the selected sets of spatial and non-spatial attribute variables yield four homogeneous geographic strata with respect to the factors that influence performance of schools. These strata clearly cut across district boundaries and carry all five types of schools.

The survey designs that use the strata constructed using spatial variables and non-spatial attribute variables are likely to yield low standard errors of the estimates as strata were constructed by controlling many variables that influence the performance. This could be verified by estimating the parameters and their standard errors. The system of classification of schools by using both spatial and non-spatial attribute variables has a special significance, as this can be used to conduct one stage stratified random sampling design to compute estimates at sub-region levels such as district, DS Division *etc.* This could be considered as a very convenient and easily implementable sampling design. This classification can also be used as a system of classifying schools according to the facilities influencing performances.

ACKNOWLEDGEMENTS

The authors are grateful to the assistance extended by the officials of the Examination Department, Surveyor General's Department, Ministry of Education, International Irrigation Management Institute and International Livestock Research Centre - Kenya.

REFERENCES

Byrne, D.S. (1959). Deindustrialization and dispossession. Sociology. 29: 95-116.

Dan, H. (2000). A procedure for stratification by an Extended Ekman Rule. J. Official Statistics. 16: 15-29.

Green, C., Morojele, N. and Maritz, J. (1998). GIS tools to bring services closer to people. A case study on planning police facility locations in Kahayelitsha, CISIR Transpoterk, South Africa.

Naude, N. (1995). Planning support tools for addressing accessibility and related rural development problems. A case study. CISIR Transpoterk, South Africa.

Solano, C., Herrero, M., Leon, H. and Perez, E. (1995). Characterizing objective profiles of Costa Rican dairy farmers. Institute of Ecology and Resource Management, The University of Edinburgh, Scotland.

Unwin, A., Unwin, D. and Fisher P. (1996). Exploratory spatial data analysis. Department of Geography, University of Leicester, United Kingdom.