USE OF GEOGRAPHIC INFORMATION SYSTEMS IN STRATIFIED SAMPLING DESIGNS. AN APPLICATION IN EDUCATION RESEARCH.

By

AMARAJEEWA JINABANDU SATHARASINGHE

Thesis

Submitted in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in the

POSTGRADUATE INSTITUTE OF AGRICULTURE

of the

UNIVERSITY OF PERADENIYA

PERADENIYA

August , 2002

Use of Geographic Information Systems in Stratified Sampling Designs - An Application in Educational Research -

ABSTRACT

In simple random sampling, the sample is not forced to be representative of different groups in the population. In stratified sampling, a sample, representative of different groups could be drawn and consequently the variance within stratum could be minimized. Therefore, in demographic and socio-economic surveys, stratified sampling designs are frequently used. Regions (geographical administrative units) are generally used as strata. Assuming causal mechanisms are localized and ignoring impacts from the bordering regions, sampling units within geographic regions are considered to be homogeneous in these surveys.

However, sampling units within geographic regions may not be homogeneous and therefore, the sample may not be representative, if the variation of spatial variables affecting the study variables were not considered in stratification. The values of the spatial variables cannot be assumed as fixed and random at every point within the geographic regions. Therefore, sampling units drawn from strata based on administrative boundaries cannot be thought to be representative of wider areas and hence do not offer scope for the generalization or extrapolation of research results.

In certain situations, use of one or two variables may not be sufficient to do a proper stratification of a population, as the population is heterogeneous due to the variations of several variables. Therefore, it may be necessary to take all such important variables in to consideration. In this study, a new stratification method is introduced to overcome these problems.

The principal used here is that the stratification is done by using sufficient number of both non-spatial variables, and spatial variables computed by using Geographic Information Systems (GIS) techniques. To be able to utilize sufficient number of variables, stratification was done by using two multivariate techniques namely factor analysis and cluster analysis, and by the Extended Ekman Rule (EER). The EER is a uni-variate stratification method. However, the use of EER was demonstrated in a multidimensional approach. The use of

linear combination of variables in reducing multistage stratified sampling designs to one-stage stratified sampling design is also demonstrated.

The new stratification method was tested by applying it to estimate parameters on the academic performance of year 5 students of public schools of the Southern Province of Sri Lanka. Academic performance was estimated by two-stage and one-stage stratified sampling designs using strata constructed under the new methods. Area was stratified using the spatial variables on the infrastructure facilities thus forming the first stage strata and schools were stratified by using non-spatial variables on the facilities available in schools and these school strata were used as second stage strata in two-stage stratified sampling designs. Schools were also stratified by spatial as well as non-spatial variables and the resulting strata were used in one stage stratified sampling design.

The one-stage stratified sampling design, using strata constructed under Extended Ekman Rule using linear combination of several factors, yielded the highest precision of the estimate. The strata constructed by this facility-based classification of schools can be used to grade schools and such a grading system can be used to determine cutoff marks for university entrance and national level examinations.

Almost all-important features of sampling such as randomness, homogeneity of strata as well as the findings can be visualized by digital thematic maps on a computer screen and can be verified against field conditions easily, by user-friendly facilities available in any common GIS software. Further, under this methodology once the system is established, samples could be drawn according to the desired requirements quickly and easily without any manual work including drawing of new geographic sub-regional boundaries and drawing sampling units according to random numbers manually.