A Geographical Study on the Aggregation Problem of Statistics:
In case of the 2000 Population and Housing Census in Korea

Gyehwa KANG

A dissertation submitted to the Doctoral Program
in Geoscience, the University of Tsukuba
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Science

January   2003

ABSTRACT

This empirical study which was not tried ever has its focus on the effectiveness of aggregation based on the individual census data. We can usually obtain aggregate data, which could harm the original characteristic of statistics as the census-related articles of the Statistics Law prevent census results from being offered to the public. We can find a poor individual in a rich community. This phenomenon is familiar to us as an ecological fallacy.

The authorities concerned about this fallacy have offered census data based on more detailed areal classification such as Mesh, small area, or Basic Unit Block(BUB), etc., on which data are processed to census users' interests. However, such data were useful only to develop a new type of statistical methods as they were published without individual nor regional characteristics.

A mobile GIS loaded with GPS enables a dynamic modifiable areal classification by creating areal coordinates. Statisticians are able to solve the so-called problem such as modifiable areal unit problem, factorial ecology and ecological fallacy.

Korea is famous for its homogeneity as it has a comparatively small territory with advanced transportation and communication system as well as one language and race. This homogeneity is not free from the aforementioned aggregation problem. This study aims to solve the aggregation problem and factorial ecology. I have created a set of 91 areal units based on 280 basic units in *Nonhyun-2 dong* to solve zoning problem and scale problem. We can easily recognize the importance of

areal classification as statistics were different according to areal classification.

This study has also ascertained the existence of factorial ecology in Korea. I have adopted a factor and cluster analysis to set up basic unit geodemographics of *Nonhyun-2 dong*. The census results were different according to selected variables and clusters to establish geodemographics, a famous statistical method. The statistics-taking authorities could reduce the common fallacy by suggesting a role model based on regular areal classification units and variables. The public can consider census results as public goods as well as private assets only after they receive a high quality census result.

This study has established a new type of census mapping system based on individual census data by utilizing mobile GIS. This newly established system is a new paradigm of managing census results focused on individual data. Furthermore, This system could serve as a naive solution to the aggregation problem such as MAUP.

Nonhyun-2 dong, Kangnamgu, Seoul, poses a vertical spatial structure as it covers high apartments area developed by the 1980s' urbanization project. Such an area comprising a variety of households could not be categorized into homogeneous areal units. Therefore, this study has proved the effectiveness of individual data by creating small area statistics based on basic, household and regional units of *Nonhyun-2 dong.* This study can greatly contribute to solve the aggregation problem now that it enables statisticians to select optimal areal classification according to their own interests.

Key words: census, aggregation problem, individual census mapping system, modifiable areal unit, Korea.