Using Principal Component Scores as Stratification Variable: An Alternative to Multiple Frame Sampling Methodology

Erniel B. Barrios
University of Philippines Diliman, Quezon City, Philippines
ernielb@yahoo.com

Kevin Carl P. Santos*
University of Philippines Diliman, Quezon City, Philippines
kevincarlsantos@gmail.com

In the Philippines, several studies pointed out that samples from Rice and Corn Production Survey no longer suffice to provide "acceptable" estimates of Livestock and Poultry statistics. Simulation studies considered the use of inventories of different animals as stratification variables or size measures. However, the results showed that the use of a particular animal type would not necessarily yield efficient estimates for other animal inventories. This leads to the use of multiple frame sampling methodology as an ideal data collection method which entails high costs. This paper proposes the use Principal Component (PC)-based scores as the stratification variable as an alternative to multiple frame sampling to lower the costs because a PC summarizes information contained in a set of auxiliary variables (e.g. inventory for different animal types); nevertheless, PC's also give premium to variables with large variability. Thus, the relatively "rare" animals, i.e. those present in very few *barangays* only, can potentially sway the PC to their advantage at the expense of the more "common" and, perhaps, the more important animal types. Hence, the authors recommend that only the inventories of the more important animals are included in the generation of the PC scores that will be used to stratify the population of interest. Various sampling experiments are performed using *barangay* level data on different animal inventories in order to determine the efficiency of the estimates using the proposed method.

Keywords: principal component analysis, livestock and poultry statistics, stratified sampling