RESEARCH REPORT ON THE
REDESIGNING OF THE BACKYARD LIVESTOCK AND POULTRY SURVEY

Jay R. Manlapaz[1], Ibarra Aaron R. Poliquit[1], Divina Gracia L. del Prado[2],
Abubakar S. Asaad[1,3], Erniel B. Barrios[4]

The main objective of the Backyard Livestock and Poultry Survey (BLPS) is to estimate the total population of major livestock and poultry coming from backyard farms. Moreover, the existing design of BLPS assumes that crop farming households are more likely to raise livestock and poultry animals that is why BLPS uses the sampling frame of Palay and Corn Production Survey (PCPS). Given the current setting, the Philippine Statistics Authority conducted a simulation study to test the assumption of the existing design and to come up with a sampling design which efficiently estimates the population of major livestock and poultry using only a single sample taken from a single sampling frame that is different and independent from the PCPS sampling frame.

Based on simulation results, the best sampling design for carabao, cattle, chicken and swine inventory estimation in terms of average CV and MAPE is a PPS systematic sample of barangays with 20 as livestock/poultry farm household sample size per sampled barangay and the best size measure variable is the sum of all livestock and poultry inventories listed in the sampling frame. Further study is still required for the case of duck and goat inventory estimation since the overall best among all explored designs during simulation still yielded high values of average CV and MAPE for all domains.

Keywords: *Lavallee-Hidiroglou Stratification Algorithm, Probability Proportional to Size (PPS) Systematic Sampling, Simulation*

I. INTRODUCTION

The Backyard Livestock and Poultry Survey (BLPS) of the Philippine Statistics Authority (PSA) is a major survey which aims to generate primary data on supply and disposition of animals from backyard farms. Information derived from BLPS should be accurate and precise in order to be an effective basis of agricultural policies, for instance, relating to local and international meat trading and farmer support programs.

The BLPS uses the sampling frame of the Palay and Corn Production Survey (PCPS) but covers only one (1) out of its four (4) replicates. The barangay sample size of a domain depends on its crop classification, that is, whether its major crop is either palay or corn. Ten (10) sample barangays are drawn when the major crop of a domain is either palay or corn while only five (5) sample barangays are drawn when the major crop is neither palay nor corn. This means that there is an assumption of significant correlation between palay/corn farming, and livestock and poultry raising in a barangay. In other words, palay/corn farming barangays are assumed to be the most likely raisers of livestock and poultry.

---

[1] Statistical Methodology Unit, Office of the National Statistician, Philippine Statistics Authority
[2] Economic Sectoral Statistics Service, Philippine Statistics Authority
[3] College of Public Health, Department of Epidemiology and Biostatistics, University of the Philippines Manila
[4] School of Statistics, University of the Philippines Diliman

However, based on barangay-level CAF 2012 data, the correlations of either palay or corn farm area to major livestock and poultry are insignificant for all domains. Table 1 below shows the correlation matrix of major livestock/poultry (number of heads) and area (in hectares) planted with palay/corn during the first and second semester of 2012 in Abra province. [*Please see Appendix for the correlation matrix of other domains.]

Table 1. Livestock/Poultry and Palay/Corn Area Correlations in Abra Province, 2012

	First Semester		Second Semester	
	Palay	**Corn**	**Palay**	**Corn**
carabao	-0.0108	0.0102	0.3036	-0.0055
cattle	0.1432	0.2598	0.1576	0.3467
chicken	0.1879	0.0935	0.3505	0.1066
duck	0.2849	0.0683	0.2219	0.03
goat	0.0813	0.1461	0.3653	0.1895
swine	0.3542	0.0292	0.303	0.0643

These insignificant correlations are consistent with the current and widespread practice of using commercially manufactured fodder instead of grain surplus or forage as sustenance to livestock and poultry.

II. OBJECTIVE

The objective of this study is to come up with a more appropriate sampling frame and design for BLPS through the use of simulations. A sampling frame that is different from the PCPS frame and a sampling design that can accurately and precisely estimate the inventories of six different animals from a single sample.

III. SCOPE AND DELIMITATION

For optimal results, the study aims to compare all plausible sampling scenarios, but the number of sampling designs that can be compared through simulation depends on the available time and computing power. This means that the number of sampling scenarios to be compared will be reduced if available time and computing power shall be lessened due to unforeseen circumstances.

IV. DESCRIPTIVE STATISTICS

There are stratification/size measure variables used in this study that were derived by simply computing the sum of heads of major livestock/poultry animals. This means that these simple sum stratification/size measure variables will perform best for animal inventories with the largest percent shares of number of heads. According to CAF 2012 data, 66.46% of total livestock/poultry number of heads in the Philippines was chicken and disaggregating by domain still puts chicken as the majority in terms of number of heads. [*Please see Appendix A for provincial distribution maps of major livestock/poultry in terms of number of heads.]

Duck and goat farms are relatively rare so compared to the other major livestock/poultry, the pairwise correlations of duck and goat inventories to all considered stratification/size measure variables are the lowest. As such, simulation results might show poor estimates for duck and goat inventories.

V. SAMPLING FRAME

The sampling frame for simulation was derived from CAF 2012 data. Household is the secondary sampling unit (SSU) of BLPS so CAF 2012 data was aggregated up to household-level, i.e. animal inventories of all operators under one household were combined. It turns out that there are 2,798,222 households out of 2,838,659 operators.

As defined by the PSA Data Archive, a backyard farm operated by a farming or non-farming farmer/household contains at least one of the following:

Livestock

1. Less than 21 heads of adult and zero head of young
2. Less than 41 heads of young animals
3. Less than 10 heads of adult and 22 heads of young

Poultry

1. Less than 500 layers, or 1000 broilers
2. Less than 100 layers and 100 broilers if raised in combination
3. Less than 100 head of duck regardless of age

From CAF 2012 data, there is no problem extracting backyard poultry farms. However, CAF 2012 data does not contain information about age of livestock so only the modified criterion – less than 41 heads – could be used to extract all backyard livestock farms. However, the modified criterion will also include small commercial livestock farms. These commercial livestock farms have less than 41 heads but fail to meet criteria 1 or 3 of backyard livestock farm classification. The trimmed household-level sampling frame has 2,783,018 households.
Lastly, all highly-urbanized cities (HUC's) were separated from respective provinces to be treated as additional domains or primary sampling units (PSU's).

VI. STRATIFICATION/SIZE MEASURE VARIABLES

Inventories of three animal groups were used to come up with all the potential stratification/size measure variables. The first group consists of carabao, cattle, chicken and swine, the second group consists of the first group plus duck and goat and the third group consists of all livestock and poultry listed in CAF 2012. These groups will be referred to as the four main, six main and all animals throughout the paper unless otherwise stated.

The stratification variables were derived from the animal groups by either taking the sum, the first principal component (scaled such that the values will range from 0 to 1) or the weighted sum where the weights per domain are computed as follows:

$$w_i = \frac{h_i l_i}{\sum_{j=1}^{a} h_j l_j}$$

where
w_i = weight for the i^th^ animal inventory
h_i = i^th^ animal inventory
l_i = i^th^ animal live weight
a = number of animal types in the animal group

Table 2 below contains the names of the potential stratification variables, the animal group from which they were based and how they were derived.

Table 2. Summary Table of Stratification/Size Measure Variables

Stratification/Size Measure Variable	Animal Group	Derivation
blps	six main animals	sum
fourmain	four main animals	sum
all_animals	all animals	sum
aggre.wtd	six main animals	weighted sum
PC1blps	six main animals	first principal component
PC1fourmain	four main animals	first principal component
PC1all	all animals	first principal component

VII. SAMPLING DESIGNS CONSIDERED

A. Sample Selection Methods

The first stage sample selection methods considered are systematic sampling and probability proportional to size systematic (PPS-Sys) sampling, where the size measure is any of the stratification variables derived. The second stage sample selection methods considered are simple random sampling without replacement (SRSWOR) and systematic sampling.

B. First Stage Sampling Designs

1. Stratification via Cumulative root frequency rule by Dalenius and Hodges (1959) – an approximation method based on constructing equal intervals on the cumulative of the square roots of the frequencies of the stratification variable so that stratum boundaries that minimize the variance are obtained.

2. Stratification via Lavallèe & Hidiroglou (LH) algorithm (1988) – an algorithm which finds stratum boundaries such that there is a take-all top stratum and a specified level of coefficient of variation and allocation scheme.

3. Stratification via deviation from the mean – since number of heads follows a Poisson distribution, compute mean and variance of stratification variable. Stratum boundaries are:

   Stratum 1: $< \bar{y} - s.d\ (y)$
   Stratum 2: $\bar{y} - s.d\ (y)\ to\ \bar{y}$
   Stratum 3: $\bar{y}\ to\ \bar{y} + s.d\ (y)$
   Stratum 4: $> \bar{y} + s.d\ (y)$, possibly take all

4. Sample selection with no stratification – select barangays via PPS-Sys (no stratification)

All stratification methods arrange the strata from lowest to highest valued in terms of the stratification variable.

The second stage (household) sample sizes considered are ten (10), fifteen (15) and twenty (20). Only five (5) stratas were considered for stratified designs. *Table 2* below shows the details on all simulation scenarios.

Table 3. Summary Table of Simulation Scenarios

SIMULATION SCENARIO	
1. Design/Stratification Method	• PPS-Sys (no stratification) • Stratified via LH algorithm • Stratified via cumulative root frequency method • Stratified via plus-minus one sd deviation from mean
2. Selection of Barangays	• PPS-Sys, Systematic (except PPS-Sys design)
3. Selection of Households	• SRSWOR, Systematic
4. No. of Sample Households	• 10, 15, 20
5. Stratification Variables	• *[See section 3 for the description of the 7 stratitification variables]*
Total: 294 scenarios = (7 stratification variables) x (42* sampling scenarios) **Note that the PPS-Sys design with no stratification has only one barangay selection method.design)*	

VIII. METHODOLOGY

A. Estimators

The design-based estimators are computed as follows:

$$\hat{Y} = \sum\nolimits_{h=1}^{L} \sum\nolimits_{i=1}^{n_h} \sum\nolimits_{j=1}^{m_{hi}} w_{hi} y_{hij}$$

$$\hat{V}(\hat{Y}) = \sum_{h=1}^{L}(1-f_h)\, n_h s_h^2 + \sum_{h=1}^{L} f_h \sum_{i=1}^{n_h}(1-f_{hi})\, m_{hi} s_{hi}^2$$

where

$\hat{Y} = DESIGN - BASED\ estimator\ of\ domain\ total$

$y_{hij} = total\ for\ SSU\ j\ in\ PSU\ i\ in\ stratum\ h$

$N_h = total\ number\ of\ PSUs\ in\ stratum\ h$

$n_h = PSU\ sample\ size\ in\ stratum\ h$

$M_{hi} = total\ number\ of\ SSUs\ in\ PSU\ i\ in\ stratum\ h$

$m_{hi} = SSU\ sample\ size\ in\ PSU\ i\ in\ stratum\ h$

$w_{hi} = w_{hij} = sampling\ weight\ for\ SSU\ j\ in\ PSU\ i\ in\ stratum\ h$

$w_{hi} = w_{hij} = \frac{N_h}{n_h} \times \frac{M_{hi}}{m_{hi}}\ for\ 1st\ stage\ systematic\ sampling$

$w_{hi} = w_{hij} = \frac{\sum_{k=1}^{N_h} X_{hk}}{n_h X_{hi}} \times \frac{M_{hi}}{m_{hi}}\ for\ 1st\ stage\ PPS - Sys\ sampling$

$\hat{V}(\hat{Y}) = DESIGN - BASED\ estimator\ of\ variance\ of\ \hat{Y}$

$s_h^2 = sample\ variance\ for\ stratum\ h\ (1st\ stage)$

$s_{hi}^2 = sample\ variance\ for\ PSU\ i\ in\ stratum\ h\ (2nd\ stage)$

$f_h = \frac{n_h}{N_h} = 1st\ stage\ sampling\ fraction\ in\ stratum\ h$

$f_{hi} = \frac{m_{hi}}{M_{hi}} = 2nd\ stage\ sampling\ fraction\ in\ stratum\ h$

Designs with no stratification is just the case when there is only one stratum (L=1) so the formulas above will still apply. Variance of estimates based on PPS systematic sampling will also be approximated using the variance formula above.

Stratification via LH-algorithm excludes stratum 1 (lowest valued) for sample selection so the estimates are computed without including h = 1 then scaled as follows:

$$SF = \left(\sum_{h=1}^{H} T_{xh}\right) \Big/ \left(\sum_{h=2}^{H} T_{xh}\right)$$
$$\hat{Y}_{LH} = SF \times \hat{Y}$$
$$\hat{V}(\hat{Y}_{LH}) = SF^2 \times \hat{V}(\hat{Y})$$

where:
$SF =$ scaling factor that is always greater than 1
T_{xh} =stratification/size measure variable total for stratum h

The average CV estimator and MAPE are computed as follows:

$$\sum_{k=1}^{R} \frac{\sqrt{\hat{V}(\hat{Y}_k)}}{\hat{Y}_k}$$

$$MAPE = \frac{1}{R}\sum_{k=1}^{R} \frac{|Y - \hat{Y}_k|}{Y}$$

where

R = number of replicates
Y = true domain total based on sampling frame
$\hat{Y}_k$ = estimator for total

B. 1st and 2nd Stage Sample Selection

The secondary sampling unit (SSU) for this study is backyard livestock/poultry farm. The steps for the 1st and 2nd stage sample selection are as follows:

1. Compute $n.annual = 2 \times n_0$ and $n = n.annual/4$ where $n.annual$ is the annual barangay sample size, n is the barangay sample size per quarter and n_0 is the barangay sample size computed using strata.LH function of R such that the stratum number is equal to 5, and the CV is equal to 0.05 for a specific stratification variable, while the constant 2 is the design effect adjustment due to cluster sampling.

2. For stratified designs, stratify the barangays via any of the considered methods.

3. Sort the barangays per stratum from lowest to highest value of stratification/size measure variable

4.
   a. Compute the following to implement PPS systematic sampling of barangays per stratum:
      - Cumulative sum of X, where X is the size measure
      - Sum of size measure of all PSUs in stratum h in the province $= \sum_{k=1}^{N_h} X_{hk}$
      - Stratum h sampling interval $K_h = (\sum_{k=1}^{N_h} X_{hk})/n_h$

   b. Or, compute the following to implement systematic random sampling of barangays per stratum:
      - Label the sorted barangays from 1 to N_h
      - Stratum h sampling interval $K_h = N_h/n_h$

5. Lastly, compute for
   - Stratum h random start from 1 to K_h
   - $R_h + (i-1) * K_h$ where $i = 1,2, \ldots n_h$

6.
   a. For PPS systematic sampling - Label a barangay as the ith sampled barangay when its corresponding cumulative size measure is less than or equal to $R_h + (i-1) * K_h$.

b. For systematic random sampling - Label a barangay as the i[th] sampled barangay when its label is equal to $R_h + (i-1) * K_h$.

7. Replace the sample duplicates by doing steps 3 and 4 for the unselected barangays and by setting sample size equal to the number of duplicates.

8. Repeat step 4 until all barangays in the sample are distinct.

9. Divide the sample into 4, one for each quarter, such that each division inherits the characteristics of a sample taken using PPS systematic sampling.

10. In each sampled barangay, select backyard livestock/poultry farm via Simple Random Sampling Without Replacement (SRSWOR).

11. Explore different backyard livestock/poultry farm sample sizes (10, 15 or 20) in each sampled barangay.

C. Trimming Down of Stratification Variables for Simulation

Based from previous simulations where there is only 1 stratification variable, it takes about 14 hours to produce output for 5 provinces. However, the simulation procedure for BLPS requires 7 stratification variables and an additional sampling design so simulation for 5 provinces will take about 14 * 7 * 1.4 = 137.2 hours which is not feasible as far as integrity of computers and working hours are concerned.

Aside from improving the R simulation codes, the possibility of reducing the number of stratification variables was also explored. The following table is the correlation matrix of stratification variables for Abra province. [*Please see Appendix for the correlation matrix of other domains.]

Table 4. Correlation Matrix of Stratification Variables (Abra province), 2012

	aggre.wtd	all_animals	blps	fourmain	PC1all	PC1blps	PC1fourmain
aggre.wtd	1.00	**0.88**	**0.88**	**0.86**	**0.81**	**0.81**	**0.81**
all_animals	**0.88**	1.00	1.00	1.00	0.98	0.98	0.98
blps	**0.88**	1.00	1.00	1.00	0.99	0.99	0.98
fourmain	**0.86**	1.00	1.00	1.00	0.99	0.99	0.99
PC1all	**0.81**	0.98	0.99	0.99	1.00	1.00	1.00
PC1blps	**0.81**	0.98	0.99	0.99	1.00	1.00	1.00
PC1fourmain	**0.81**	0.98	0.98	0.99	1.00	1.00	1.00

It is apparent from the correlation matrix that all pairwise correlations have values close to 1 so results using all stratification variables are expected to be similar. However, pairwise correlations of aggre.wtd with the rest are the lowest. This is also the case for all the other domains. Hence, this will be the basis for reducing the number of stratification variables to 2

– aggre.wtd and all_animals since it is one of the easiest to compute among the remaining stratification variables. Since only 2 stratification variables will be used for phase 2 simulations, running time for 5 provinces will be reduced to 137.2 * (2 / 7) = 39.2 hours.

D. Simulation

Simulation was done using R such that the goal is to select the best sampling design. The first simulation involved taking one hundred (100) sample barangay replicates and one hundred (100) sample farm replicates per sampled barangay replicate for a total of ten thousand (10, 000) barangay-farm sample replicates. The average CV as precision measure and the mean absolute percentage error (MAPE) as accuracy measure per sampling scenario were computed using the ten thousand (10, 000) barangay-farm sample replicates. The criteria for selecting the best design are design simplicity, average CV and MAPE.

IX. RESULT AND DISCUSSION

All sampling scenarios were evaluated in terms of average CV and MAPE of the design-based estimates. The average CV measures precision while the MAPE measures accuracy.

All sampling scenarios are designed to have independent quarterly estimates. However, the sampling frame does not disaggregate or classify animal inventories according to quarter which means that results per quarter are expected to be similar. Indeed, in turned out that across animal inventories, only one sampling design is best for all quarters.

The simulation results indicate that for all provinces and animals and in terms of lowest average CV and MAPE, the best barangay selection method is PPS-Sys, the best farm selection method is SRSWOR and the best farm sample size is twenty (20).

There is no uniform best stratification/size measure variable and sampling design, but the list were narrowed down to two (2) sampling designs, namely, stratification via LH-algorithm and PPS-Sys with no stratification. The following table show the best design per main animal inventory in terms of average CV or MAPE.

Table 4.

Measure	Animal	Best Design	Best Stratification/Size Measure Variable
Average CV	Carabao	Stratified LH	aggre.wtd
	Cattle	Stratified LH	aggre.wtd
	Chicken	Stratified LH	all_animals
	Duck	PPS-Sys	all_animals
	Goat	Stratified LH	all_animals
	Swine	Stratified LH	aggre.wtd
	Overall	Stratified LH	all_animals
MAPE	Carabao	PPS-Sys	aggre.wtd
	Cattle	PPS-Sys	aggre.wtd
	Chicken	PPS-Sys	all_animals
	Duck	PPS-Sys	all_animals
	Goat	PPS-Sys	all_animals
	Swine	PPS-Sys	aggre.wtd
	Overall	PPS-Sys	all_animals

The overall best stratification/size measure variable is all_animals. The best design in terms of MAPE is PPS-Sys and the best design in terms of average CV is Stratified LH. Excluding the extracted HUC's, the differences in average CV between Stratified LH and PPS-Sys is close to zero but considering that PPS-Sys is a lot more simpler than Stratified LH, then the overall best design is PPS-Sys.

The following table gives the summary of the best sampling scenario according to average CV, MAPE and simplicity.

Table 5.

Best 1st Stage Sampling Design	PPS-Sys with no stratification
Best Barangay Selection Method	PPS-Sys
Best Farm Selection Method	SRSWOR
Best Farm Sample Size	20
Best Size Measure Variable	all_animals

X. CONCLUSION

Based from the simulation results, the best first stage sampling design is PPS systematic sampling with no stratification, the best livestock/poultry farm household selection method is SRSWOR, the best farm sample size among those explored is twenty (20), the best size measure variable is the sum of all animal inventories listed in the sampling frame, and the best estimation procedure is 2nd stage bootstrap estimation. However, the accuracy and precision of estimates are only acceptable for the original BLPS domains (excluding extracted HUC's) and for the carabao, cattle, chicken and swine inventories (excluding duck and goat).

XI. RECOMMENDATION

Other size measure variables or even sampling designs can be explored since neither aggre.wtd nor all_animals as size measure for PPS systematic sampling performs well for all the original BLPS domains. Estimation of duck and goat inventories should be treated differently since both animals are relatively rare compared to the other BLPS animals. Lastly, estimates were unreliable for the extracted HUC’s so it is not recommended for these HUC’s to be treated as separate domains.

REFERENCES

Baillargeon, S. and L.P. Rivest (2014). Univariate Stratification for Survey Population. https://cran.r-project.org/web/packages/stratification/stratification.pdf Date accessed: November 10, 2016.

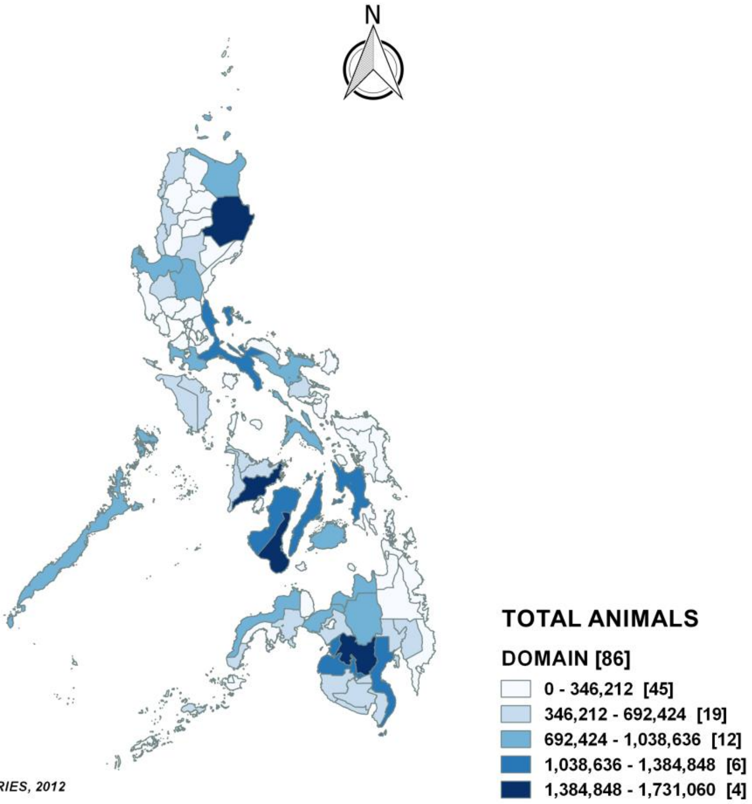
Lavallee, P. and Hidiroglou, M. (1988). On the Stratification of Skewed Populations. Sruvey Methodology. 14(1), p. 33-34.

Lohr, S. L., (2010). Sampling: Design and Analytics, Second Edition. Arizona State University, USA.

APPENDIX

Appendix A

Figure 1. Distribution of Total Animals in the Philippines, 2012



SOURCE: CENSUS OF AGRICULTURE AND FISHERIES, 2012

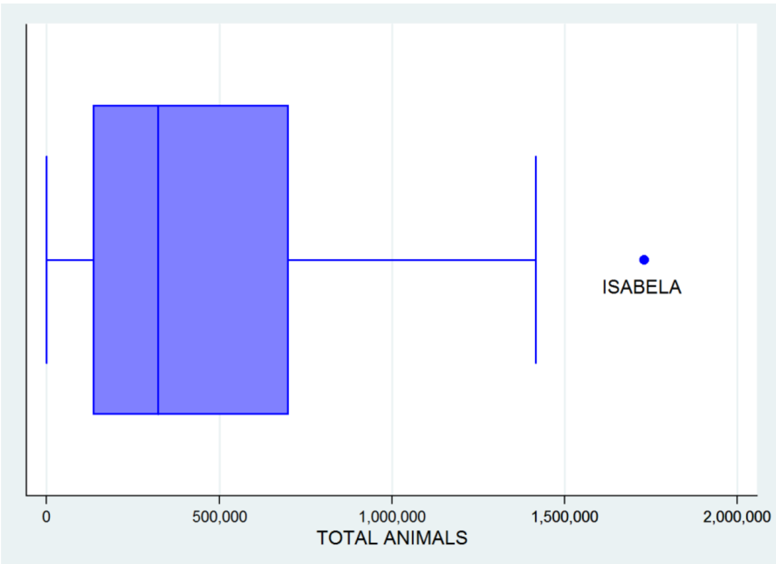
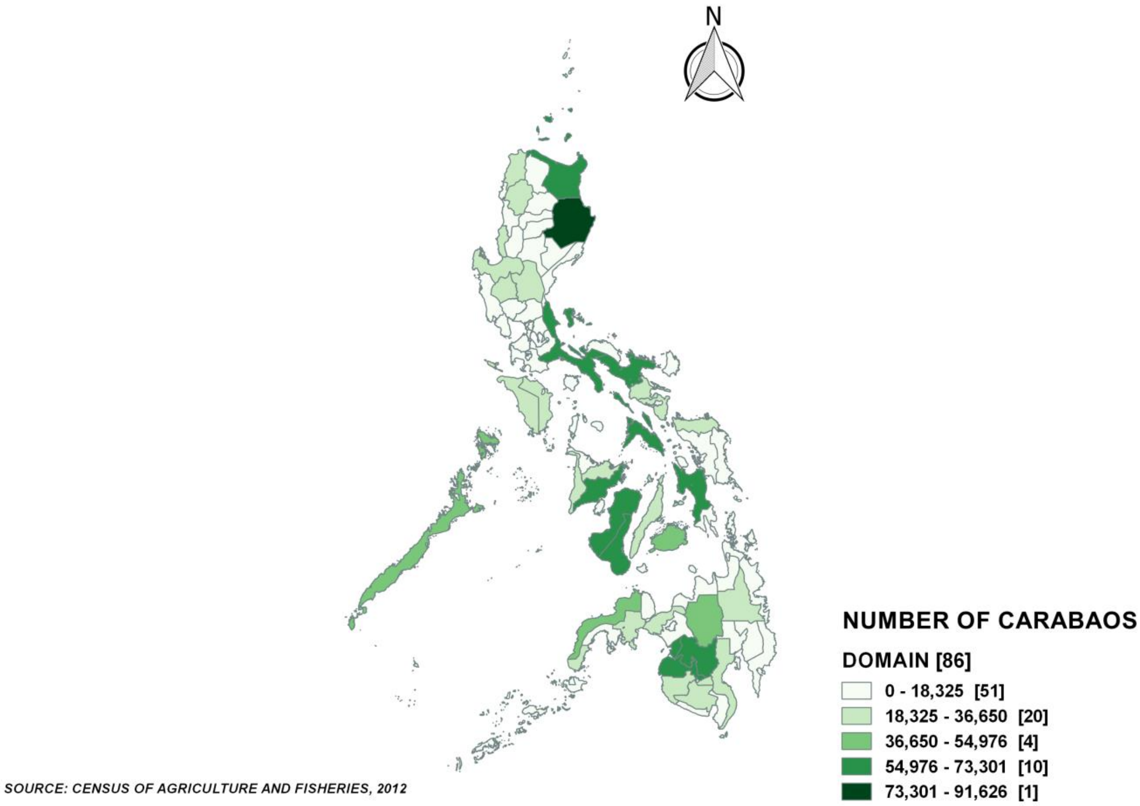


Figure 2. Distribution of Number of Carabaos in the Philippines, 2012



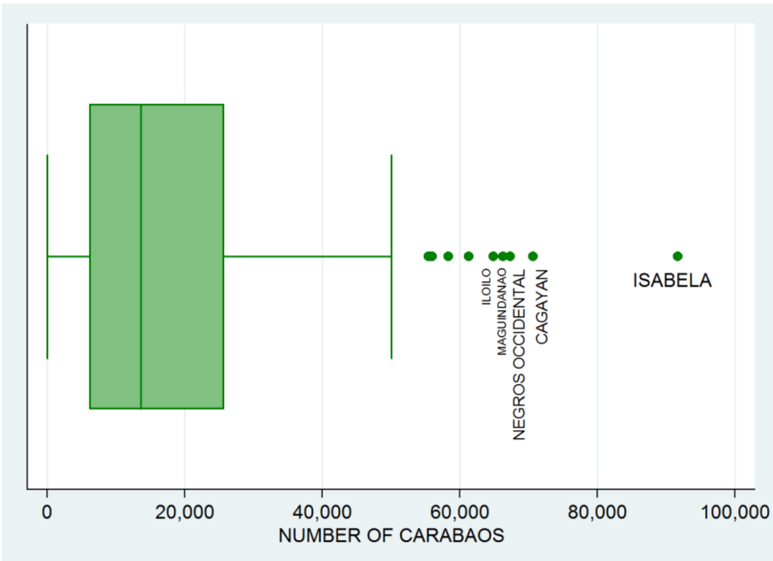
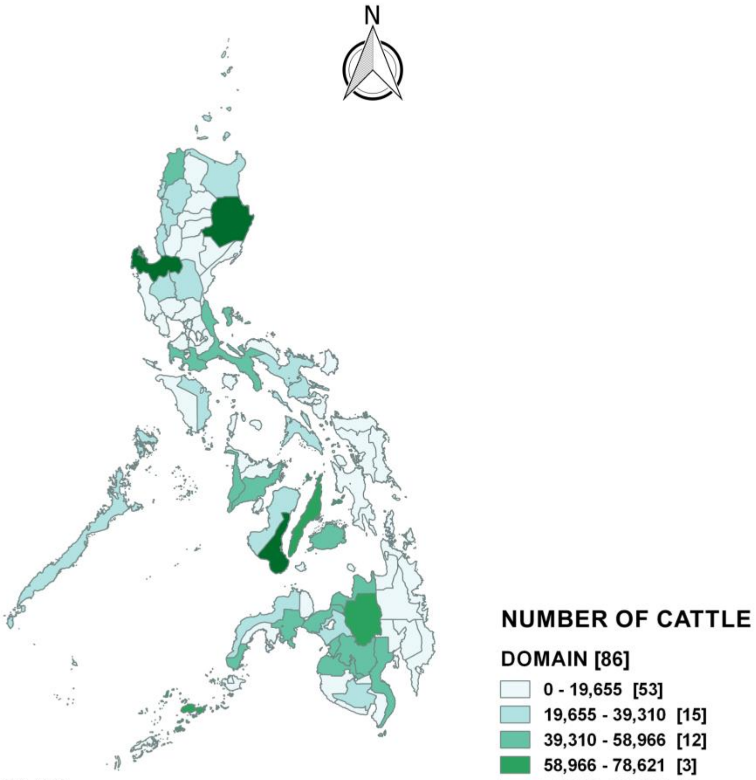


Figure 3. Distribution of Number of Cattle in the Philippines, 2012



SOURCE: CENSUS OF AGRICULTURE AND FISHERIES, 2012

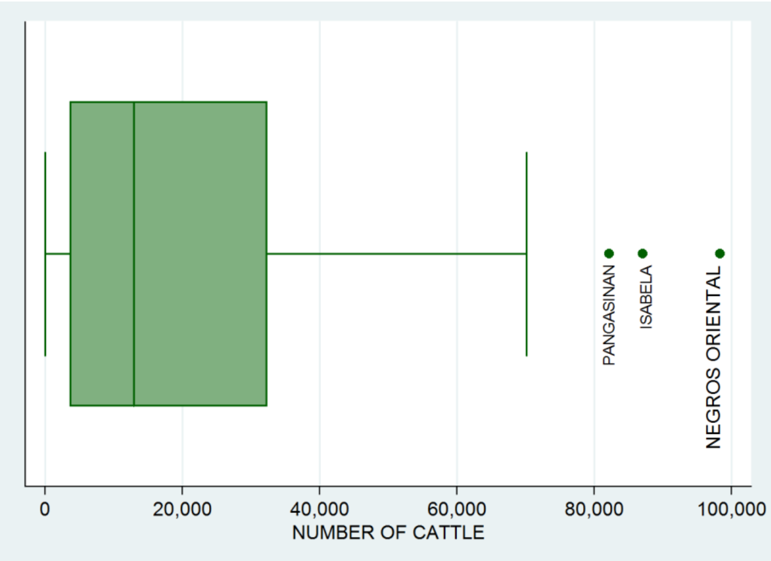
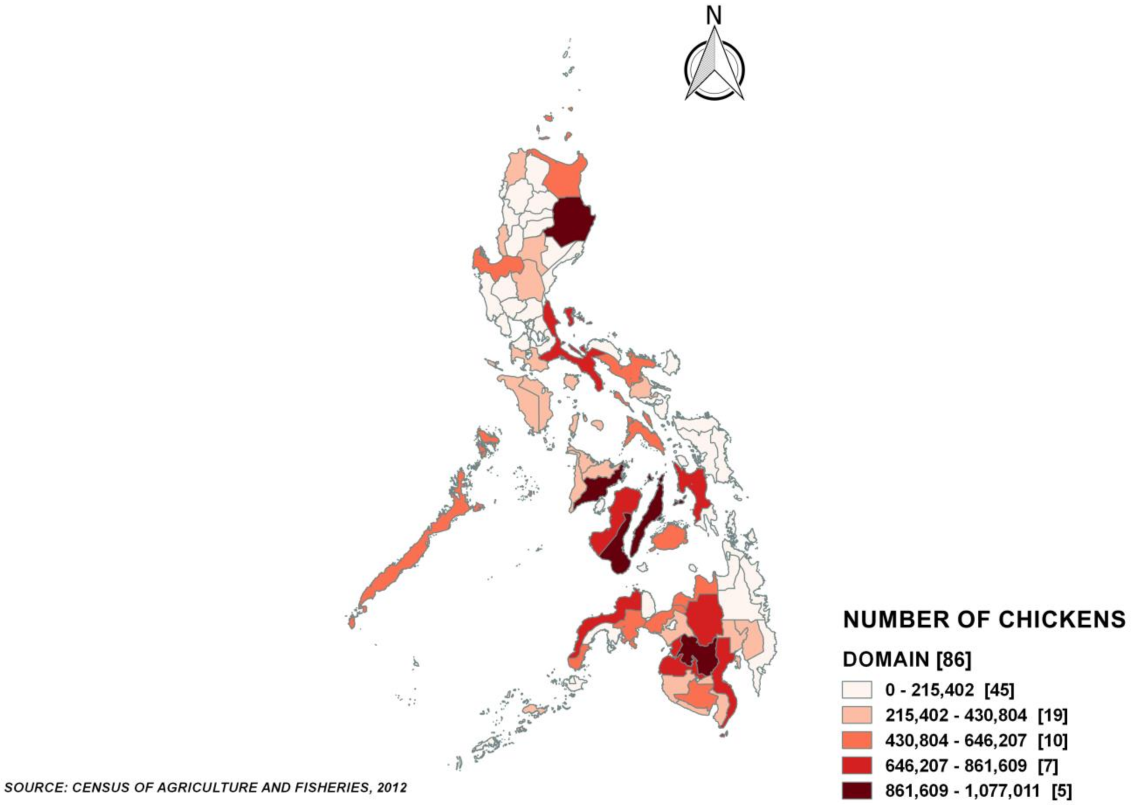


Figure 4. Distribution of Number of Chickens in the Philippines, 2012



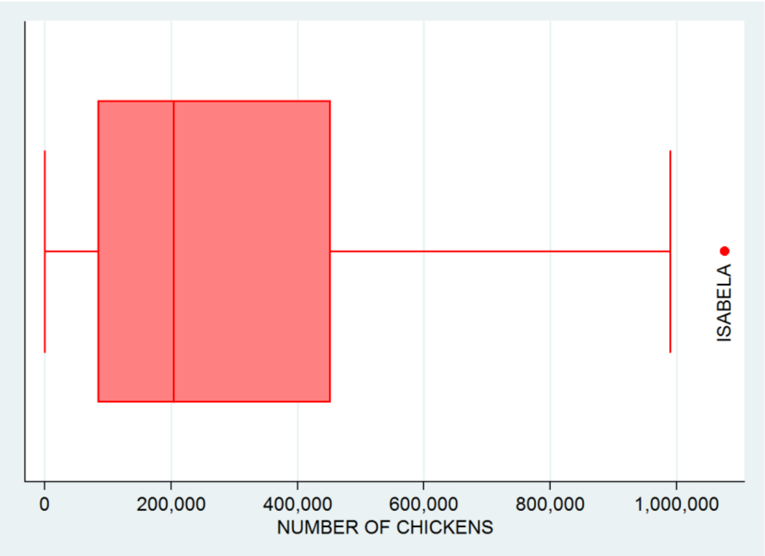
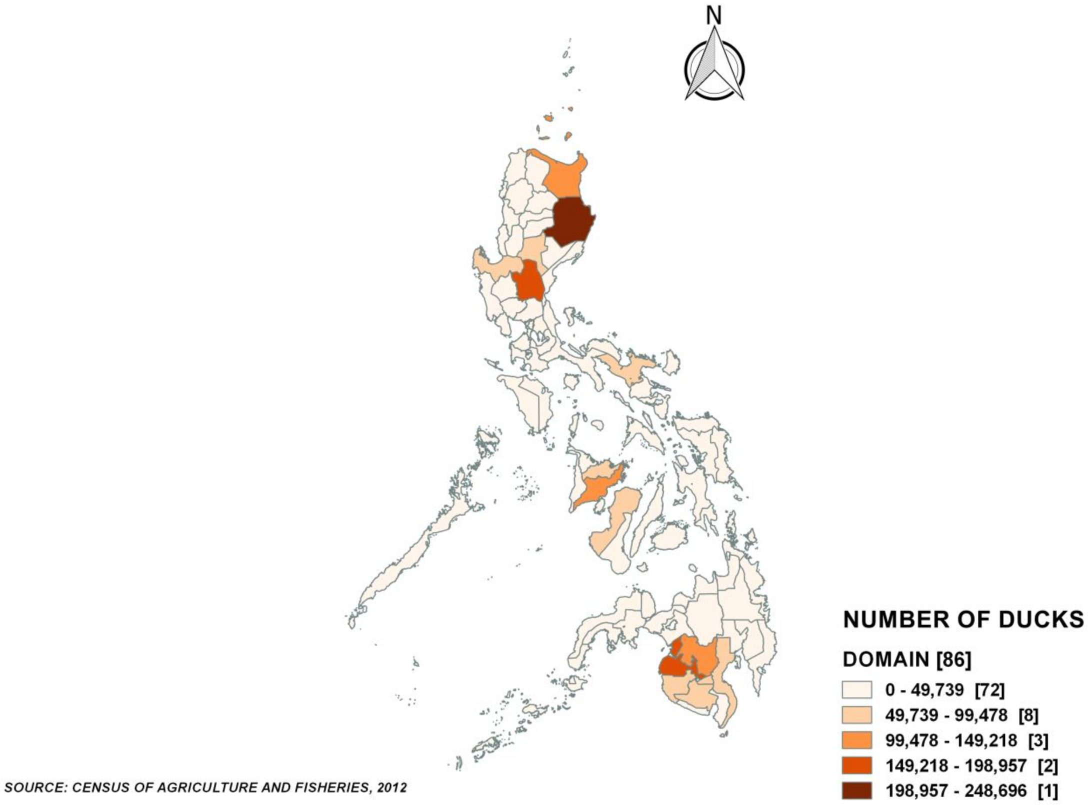


Figure 5. Distribution of Number of Ducks in the Philippines, 2012



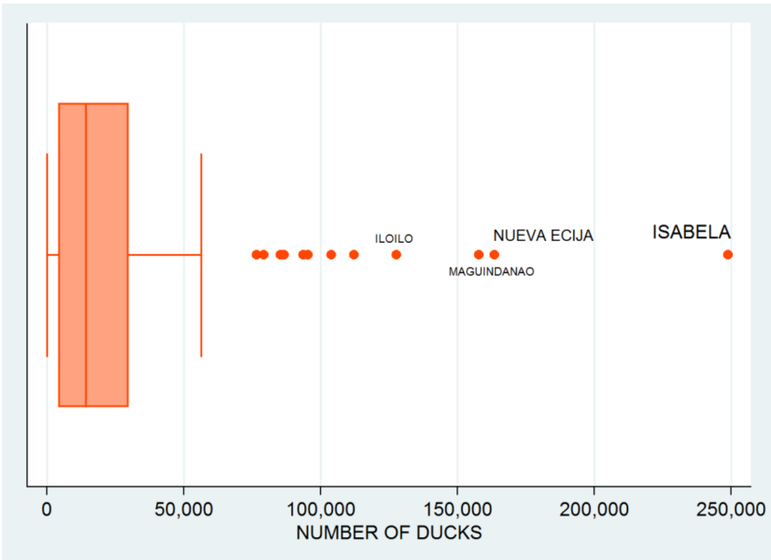
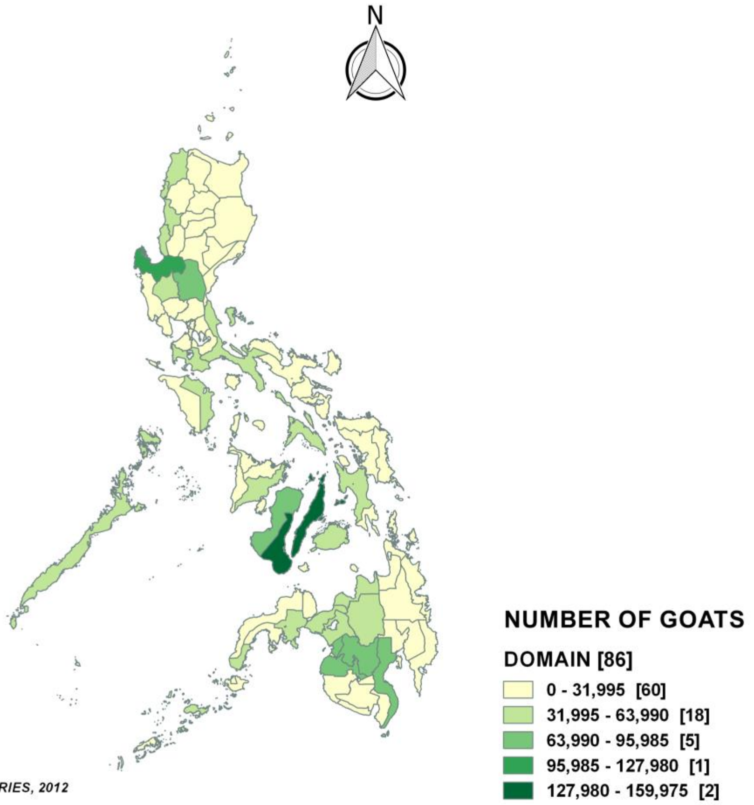


Figure 6. Distribution of Number of Goats in the Philippines, 2012



SOURCE: CENSUS OF AGRICULTURE AND FISHERIES, 2012

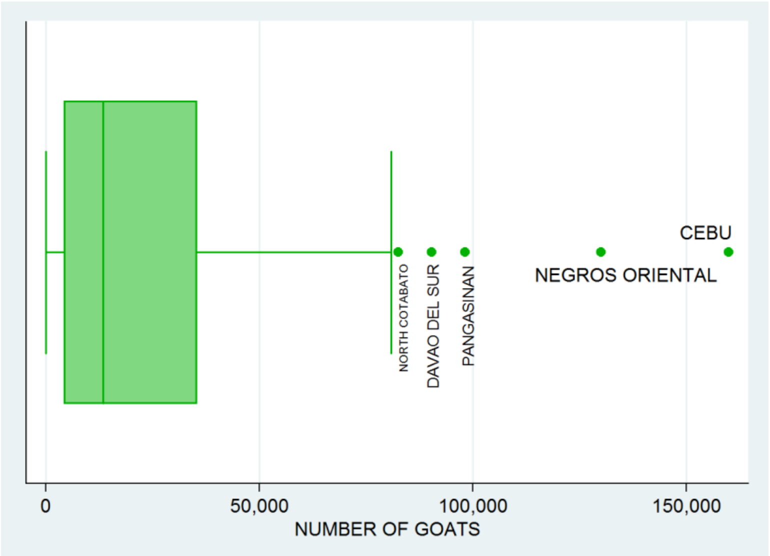


Figure 7. Distribution of Number of Swines in the Philippines, 2012