Optimizing Forest Sampling by using Lagrange Multipliers

Suhud Wahyudi, Farida Agustini Widjajati and Dea Oktavianti

***Abstract*—To obtain information from a population, we use a sampling method. One of sampling techniques that we can use is double sampling. Double sampling is a sampling technique based on the information of first phase which is used as an additional information obtaining estimates for the second phase. In this case, we discuss the model of double sampling with regression estimator. Then, to obtain the optimal number of samples for the first and second phases, we use Lagrange multipliers. The model analysis result is a formula to calculate the optimal number of samples for the first phase (n') and the second phase (n). Implementation of this method is simulated by using teak stands data from previous studies at Forest Management Unit (FMU) Madiun which consists of Section Forest Management Units (FSMU) Dagangan and Dungus. The calculation result of data from FSMU Dagangan, we get optimal number of plots must be observed in image interpretation are 149 plots and field survey are 14 plots. And with the data from FSMU Dungus, we get optimal number of plots to be observed in image interpretation are 153 plots and field survey are 20 plots.**

***Index Terms*—Double sampling, Lagrange multiplier optimizations, regression estimator.**

I. Introduction

STATISTICAL method is one branch of mathematical science that focuses on the data collection technique, processing or analyzing data, and deduction based on the data. In data processing, we analyze the relationship between two or more variables, and decide which one is the most important. To analyze the data we can use regression and correlation, in order to determine which variables are interconnected. One of the discussion in statistical methods is sampling technique. In statistical inference, if we want to obtain conclusions about the population, although without comprehensive observation, the composition of individuals in the population, we can use sampling technique [1].

We use sampling technique because of time and cost efficiency, large enough population, the precision in the execution of observation, and the value of benefits. In the process, sampling has many techniques that we can use in various implementations of sampling, one of them is double sampling. Double sampling is a sampling technique based on the information of first phase which is used as an additional information obtaining estimates for the second phase. One implementation of double sampling is in forest inventory [2].

However, we have to consider the cost factor of sampling, so that we need an optimal allocation between the number of samples in the first phase and second phase. To determine the optimal number of samples, we can use optimization by minimizing the cost function and define the variance estimator function as constraint. The result of optimization is an optimal number of samples for the first phase and the second phase [2].

We can use the Lagrange multipliers method to optimization. Lagrange multiplier method (or Lagrange multipliers) are introduced by Joseph Louis de Lagrange (1736-1813). Lagrange multiplier method is a method to maximize or minimize a function of several variables by using λ as its Lagrange multipliers. The extension of the method to a general problem of n variables with m constraints has been discussed in [3]. Kitikidou explains that sampling optimization by using Lagrange multipliers are computed by minimizing the cost function and defining variance estimator function as constraints [3]. In this paper, we discuss a sampling optimization using Lagrange multipliers method and apply it to the forest observation, especially the teak forest inventory.

II. Double Sampling, Linear Regression Model and Regression Estimator in Double Sampling

A. Two Phase Sampling (Double Sampling)

Double sampling is one of sampling techniques with two phases. In the first phase, we choose n' units number of samples, and in the second phase we choose n units that are part of the first phase. We use the first phase as an estimator for the second phase. In this case, we use regression estimator.

Mean of regression estimator is [3]:

$$\hat{\bar{y}} = \bar{y} + b(\bar{x}' - \bar{x})$$

where

- $\bar{y}$: mean of y from sub sample (n)
- $\bar{x}'$: mean of x from sample (n')
- $\bar{x}$: mean of x from sub sample (n)
- b : estimator of β

Variance of regression estimator is [3]:

$$Var\left(\hat{\bar{y}}\right) = S_y^2\left(\frac{1}{n} - r^2\frac{n'-n}{nn'}\right)$$

where,

- S_y^2 : variance of y on subsample (n)
- r : correlation coefficient between y and x
- n' : the number of first sample which is taken from N
- n : the number of subsamples from n'

Optimum allocation from cost function is [3]:

$$C = n'C_1 + nC_2$$

Manuscript received March 6, 2017; accepted May 4, 2017.
The authors are with the Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia. Email: suhud@matematika.its.ac.id

where,

- C : the total cost of sampling
- C_1 : the cost of first phase sampling
- C_2 : the cost of second phase sampling

Let $x_1, x_2, \ldots, x_n$ are n random samples from population with mean μ dan standard deviation σ. If we use sampling with replacement and unlimited population then we get [1]:

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

where,

- $\mu_{\bar{x}}$: mean of mean sampling distribution
- $\sigma_{\bar{x}}^2$: variance of mean sampling distribution

In *double sampling*, we assume $\bar{y}$ have normal distribution, confidence interval for mean is [3]:

$$\hat{\bar{y}} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{y} < \hat{\bar{y}} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $Var\left(\hat{\bar{y}}\right) = \frac{\sigma^2}{n}$, error of estimation is [3]:

$$\varepsilon = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

then we get $Var\left(\hat{\bar{y}}\right) = \frac{\varepsilon^2}{Z_{\alpha/2}^2}$.

B. Linear Regression Model

Linear regression model of population is [4]:

$$Y = \alpha + \beta X + e_j$$

where α and β are constant population parameters, and β is the regression coefficient. Regression coefficient β is [5]:

$$\beta = \frac{\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{N}(x_i - \overline{X})^2} = \frac{\sigma_{xy}}{\sigma_x^2}$$

Variance of population in regression is [5]:

$$\sigma^2 = \sigma_y^2 - \beta^2\sigma_x^2 \tag{1}$$

Correlation coefficient of population in regression is [5]:

$$\rho = \frac{\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(y_i - \overline{Y})^2}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

Relation of correlation coefficient with regression coefficient is [5]:

$$\rho = \beta\frac{\sigma_x}{\sigma_y}$$

From equation (1), we can write:

$$\sigma^2 = \sigma_y^2(1 - \rho^2) \tag{2}$$

From equation (2), we get:

$$\sigma_y^2 = \frac{\sigma^2}{(1 - \rho^2)}$$

Regression model in sample is [4]:

$$y = a + bx_k + e_k$$

For $k = 1, 2, 3, \ldots, n$ with a and b are estimators for α and β, and e_k is error of estimator for k-th observation.

Regression coefficient b is [5]:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_x^2}$$

Correlation coefficient of sample in regression is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}}{S_xS_y} \tag{3}$$

where,

$$S_y^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n - 1}$$

C. Regression Estimator in Double Sampling

Another model of Y is [5]:

$$Y = \overline{Y} + \beta(x - \overline{X})$$

If we assume y is an estimator of equation Y, then we get :

$$y = \overline{Y} + \beta\left(x - \overline{X}\right) + e$$

where e is the error, so $E(e) = 0$, then we get:

$$\bar{y} = \overline{Y} + \beta\left(\bar{x} - \overline{X}\right) + \bar{e} \tag{4}$$

We also get:

$$E(\bar{y}) = E\left(\overline{Y} + \beta\left(\bar{x} - \overline{X}\right) + \bar{e}\right) = \overline{Y}$$

The above equation of $E(\bar{y})$ shows that y is an unbiased estimator for Y.

If we assume $\frac{\sum_{i=1}^{n} e_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \bar{e}_w$, we get the value of b which is an estimator of β as follows:

$$b = \beta + \bar{e}_w$$

Because of $E(\bar{e}) = 0$, so $E(\bar{e}_w) = 0$. Then $E(b) = \beta$ and $E(E(b)) = \beta$

$$E\left(\bar{e}_w{}^2\right) = E\left(\left(\frac{\sum_{i=1}^{n} e_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^2\right) = \frac{\left(\bar{x} - \overline{X}\right)^2}{\sum\left(\bar{x} - \overline{X}\right)^2}$$

D. Expectation and Variance in Multivariate Distribution

General variance distribution is defined as [5]:

$$Var(y) = E(y - E(y))^2 = E\left(y^2\right) - (E(y))^2$$
$$E\left(y^2\right) = Var(y) + (E(y))^2$$

Expectation and variance in multivariate distribution is [5]:

$$E = E_1(E_2 \ldots (E_m))$$
$$Var = E_1(E_2 \ldots (E_m(Var_m)) + E_1(E_2 \ldots (Var_{m-1}(E_m))) + Var_1(E_2(E_{m-1}(E)))$$

E. Fisher's Distribution

Fisher's variable $F(m_1,m_2)$ is distributed as [5]:

$$F(m_1,m_2)=\frac{1}{F(m_2,m_1)}$$

where m_1 and m_2 is degree of freedomin Fisher's distribution.

If we assume $m_1=1$ and $m_2=n-1$, we get:

$$F\left(1,n-1\right)=n(n-1)\frac{\left(\overline{x}_i-\overline{X}\right)^2}{\sum_{i=1}^{n}\left(\overline{x}_i-\overline{X}\right)^2}$$

Expectation of distribution is as follows [5]:

$$E(F\left(m_1,m_2\right)=\frac{m_2}{m_2-2}$$

where $m_2\geq 3$.

F. Optimization by using Lagrange Multipliers Method

Optimization technique of multivariables with equality constraint have the following general form [6]:

$$\begin{aligned}&\text{minimize } f(X)\\&\text{subject to } g_j(X)=0,\ \text{for } j=1,2,\dots,m\\&\text{where } X=\{x_1,x_2,\dots,x_n\}^T\end{aligned}$$

where $m\leq n$. If $m>n$, then it cannot be solved.

The first step of this method is the construction of Lagrange function that is defined as [6]:

$$L(X,\lambda)=f(X)+\sum_{j=1}^{m}\lambda_j g_j(X) \tag{5}$$

Theorem 1 ([6]): Necessary condition for a function $f(X)$ with constraint $g_j(X)=0$, where $j=1,2,\dots,m$ such that it has relative minimum at point x^* is first partial derivative of Lagrange function defined as $L=L\{x_1,x_2,\dots,x_n,\lambda_1,\lambda_2,\dots,\lambda_n\}$ has value zero. ■

Theorem 2 ([6]): A sufficient condition for $f(X)$ to have relative minimum (or maximum) at the quadratic, Q, defined by:

$$Q=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 L}{\partial x_i\partial x_j}dx_i dx_j$$

evaluated at $x=x^*$ must be positive definite (or negative definite) for all values of dx for which the constraints are satisfied. ■

Necessary condition $Q=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 L}{\partial x_i\partial x_j}dx_i dx_j$ to be positive (or negative) definite for all admissible variations dx is that each root of the polynomial p_i, defined by the following determinant equation, be positive (or negative).

$$\begin{vmatrix}L_{11}-p & L_{12} & L_{13} & \dots & L_{1n} & g_{11} & g_{21} & \dots & g_{m1}\\ L_{21} & L_{22}-p & L_{23} & \dots & L_{2n} & g_{12} & g_{22} & \dots & g_{2n}\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ L_{n1} & L_{n2} & L_{n3} & \dots & L_{nm}-p & g_{m1} & g_{m2} & \dots & g_{mn}\\ g_{11} & g_{12} & g_{13} & \dots & g_{1n} & 0 & 0 & \dots & 0\\ g_{21} & g_{22} & g_{23} & \dots & g_{2n} & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ g_{m1} & g_{m2} & g_{m3} & \dots & g_{mn} & 0 & 0 & \dots & 0\end{vmatrix}=0 \tag{6}$$

where $L_{ij}=\frac{\partial^2 L(x^*,\lambda)}{\partial x_i\partial x_j}$ and $g_{ij}=\frac{\partial g_i(x^*)}{\partial x_j}$ Observe that equation (6) is a polynomial of order $n-m$, in p.

III. Results and Discussions

A. Mean of Regression Estimator in Double Sampling

Linear regression estimator can be defined as:

$$\begin{aligned}Y&=\alpha+\beta X\\ \overline{Y}&=\alpha+\beta\overline{X}\end{aligned} \tag{7}$$

Equation (7) is linear regression population average. If we estimate linear regression from its sample, where a is an estimator for α, and b is an estimator for β, so we obtain:

$$y=a+bx \tag{8}$$

$$\overline{y}=a+b\overline{x} \tag{9}$$

Equation (9) is a linear regression average equation in sample. From equation (9), we obtain:

$$a=\overline{y}-b\overline{x} \tag{10}$$

By substituting equation (10) to equation (8), we obtain estimator regression equation as follows:

$$\begin{aligned}y&=\overline{y}+bx-b\overline{x}\\&=\overline{y}+b(x-\overline{x})\end{aligned}$$

If the value of x is unknown, then to compute its estimator, we can use $\overline{x}'=\sum_{i=1}^{n}\frac{x_i}{n'}$. we obtain:

$$\widehat{\overline{y}}=\overline{y}+b(\overline{x}'-\overline{x}) \tag{11}$$

Equation (11) is mean estimator equation of linear regression in *double sampling*.

B. Variance of Regression Estimator in Double Sampling

After we obtain equation (11), substituting $\overline{y}$ from equation (4) and value of b from equation (11), so we obtain another model of mean estimator equation of linear regression in *double sampling* . It is given by:

$$\begin{aligned}\widehat{\overline{y}}&=\overline{Y}+\beta\left(\overline{x}-\overline{X}\right)+\overline{e}+(\beta+\overline{e}_w)(\overline{x}'-\overline{x})\\&=\overline{Y}+\beta\left(\overline{x}'-\overline{X}\right)+\overline{e}_w\left(\overline{x}'-\overline{X}\right)-\overline{e}_w\left(\overline{x}-\overline{X}\right)+\overline{e}\end{aligned}$$

To obtain the variance of regression estimator, we use the trivariate distribution theory. Mean of trivariate distribution is as follows:

$$E\left(\widehat{\overline{Y}}\right)=E_1\left(E_2\left(E_3\left(\widehat{\overline{y}}\right)\right)\right)$$

We assume:

$$E_3\left(\widehat{\overline{y}}\right)=\overline{Y}+\beta\left(\overline{x}'-\overline{X}\right)$$

So we obtain $E_3\left(\widehat{\overline{y}}\right)$, which is an unbiased estimator. Then we determine $E_2\left(E_3\left(\widehat{\overline{y}}\right)\right)$ as follows:

$$\begin{aligned}E_2\left(E_3\left(\widehat{\overline{y}}\right)\right)&=E_2\left(\overline{Y}+\beta\left(\overline{x}'-\overline{X}\right)\right)\\&=\overline{Y}+\beta\left(\overline{x}'-\overline{X}\right)\end{aligned}$$

For the next step, we assume $\overline{x}'$ is not constant, we obtain:

$$\begin{aligned}E_1\left(E_2\left(E_3\left(\widehat{\overline{y}}\right)\right)\right)&=E_1\left(\overline{Y}+\beta\left(\overline{x}'-\overline{X}\right)\right)\\&=\overline{Y}\end{aligned} \tag{12}$$

Equation (12) is an unbiased estimator of the trivariate distribution.

Then, we get variance of trivariate distribution by the formula :

$$Var\left(\widehat{\overline{Y}}\right) = E_1\left(E_2\left(Var_3\left(\widehat{\overline{y}}\right)\right)\right) + E_1\left(Var_2\left(E_3\left(\widehat{\overline{y}}\right)\right)\right) + Var_1\left(E_2\left(E_3\left(\widehat{\overline{y}}\right)\right)\right) \tag{13}$$

Then we determine the second part of equation (13), so we obtain:

$$E_1\left(Var_2\left(E_3\left(\widehat{\overline{y}}\right)\right)\right) = E_1\left(Var_2\left(\overline{Y} + \beta\left(\overline{x}' - \overline{X}\right)\right)\right) = 0$$

Third part of equation (13) is:

$$\begin{aligned} Var_1\left(E_2\left(E_3\left(\widehat{\overline{y}}\right)\right)\right) &= Var_1\left(\overline{Y} + \beta\left(\overline{x}' - \overline{X}\right)\right) \\ &= \beta^2 E_1\left(\overline{x}' - E_1(\overline{x}')\right)^2 = \beta^2 \frac{\sigma_x^2}{n'} \end{aligned}$$

Next, we determine:

$$\begin{aligned} &E_1\left(E_2\left(Var_3\left(\widehat{\overline{y}}\right)\right)\right) \\ &= E_1\left(E_2\left(Var_3\left(\overline{Y} + \beta\left(\overline{x}' - \overline{X}\right) + \overline{e}_w\left(\overline{x}' - \overline{X}\right) - \overline{e}_w\left(\overline{x} - \overline{X}\right) + \overline{e}\right)\right)\right) \\ &= E_1\left(E_2\left(\overline{e} - \overline{e}_w\left(\overline{x} - \overline{X}\right) + \overline{e}_w\left(\overline{x}' - \overline{X}\right)\right)^2\right) \end{aligned} \tag{14}$$

So that equation (14) becomes:

$$\begin{aligned} &= E_1\left(E_2\left(\overline{e}^2\right)\right) + E_1\left(\left(\overline{x} - \overline{X}\right)^2 E_2\left(\overline{e}_w{}^2\right)\right) + \\ &E_1\left(\left(\overline{x}' - \overline{X}\right)^2 E_2\left(\overline{e}_w{}^2\right)\right) - E_1\left(2\left(\overline{x} - \overline{X}\right) E_2\left(\overline{e}\overline{e}_w\right)\right) + \\ &E_1\left(2\left(\overline{x}' - \overline{X}\right) E_2\left(\overline{e}\overline{e}_w\right)\right) - E_1\left(2\left(\overline{x} - \overline{X}\right)\left(\overline{x}' - \overline{X}\right) E_2\left(\overline{e}_w{}^2\right)\right) \end{aligned} \tag{15}$$

Then, we solve each part of equation (15), the first part of equation (15) is:

$$E_1\left(E_2\left(\overline{e}^2\right)\right) = E_1\left(\overline{e} - E_2\left(\overline{e}\right)\right)^2 = \sigma_{\overline{e}}^2 = \frac{\sigma^2}{n}$$

Next, we solve the second part of equation (15) as follows:

$$E_1\left(\left(\overline{x} - \overline{X}\right)^2 E_2\left(\overline{e}_w{}^2\right)\right) = \frac{\sigma^2}{n^2(n-1)} E_1 F\left(1, n-1\right) = \frac{\sigma^2}{n^2(n-3)}$$

Then, we solve the third part of equation (15) as follows:

$$E_1\left(\left(\overline{x}' - \overline{X}\right)^2 E_2\left(\overline{e}_w{}^2\right)\right) = \frac{\sigma^2}{n'^2(n-1)} E_1 F\left(1, n-1\right) = \frac{\sigma^2}{n'^2(n-3)}$$

The fourth part of equation (15) is as follows:

$$E_1\left(2\left(\overline{x} - \overline{X}\right) E_2\left(\overline{e}\overline{e}_w\right)\right) = 2E_1\left(\overline{x} - \overline{X}\right) E_1\left(0\right) = 0$$

Next, the fifth part of equation (15) is given by:

$$E_1\left(2\left(\overline{x}' - \overline{X}\right) E_2\left(\overline{e}\overline{e}_w\right)\right) = 2E_1\left(\overline{x}' - \overline{X}\right) E_1\left(0\right) = 0$$

From the sixth part of equation (15), we obtain:

$$\begin{aligned} &E_1\left(2\left(\overline{x} - \overline{X}\right)\left(\overline{x}' - \overline{X}\right) E_2\left(\overline{e}_w{}^2\right)\right) \\ &= 2E_1\left(\overline{X} - \overline{X}\right) E_1\left(\overline{X} - \overline{X}\right) E_2\left(\overline{e}_w{}^2\right) \\ &= 0 \end{aligned}$$

The result from all previous steps is:

$$E_1\left(E_2\left(Var_3\left(\widehat{\overline{y}}\right)\right)\right) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n^2(n-3)} + \frac{\sigma^2}{n'^2(n-3)} \tag{16}$$

If the number of sample n is too big, then $\frac{1}{(n-3)} \approx \frac{1}{n}$ and equation (16) becomes:

$$= \frac{\sigma^2}{n}\left(1 + \frac{1}{n^2} + \frac{1}{n'^2}\right)$$

From analysis result, equation (13) becomes:

$$Var\left(\widehat{\overline{Y}}\right) = \frac{\sigma^2}{n}\left(1 + \frac{1}{n^2} + \frac{1}{n'^2}\right) + \beta^2 \frac{\sigma_x^2}{n'} \tag{17}$$

If $n' \to \infty$ and $n \to \infty$, then $\frac{1}{n} \to 0$ and $\frac{1}{n'} \to 0$. In this case, equation (17) becomes:

$$Var\left(\widehat{\overline{Y}}\right) = \sigma_y^2\left(\frac{1}{n} - \rho^2\left(\frac{n' - n}{nn'}\right)\right) \tag{18}$$

Equation (18) is a variance equation of regression estimator in population. If we estimate equation (18) in sample, then we can write:

$$Var\left(\widehat{\overline{y}}\right) = S_y^2\left(\frac{1}{n} - r^2 \frac{n' - n}{nn'}\right) \tag{19}$$

Equation (19) is variance equation of regression estimator in sample.

C. Sampling Optimization by using Lagrange Multipliers Method

First step of optimization is determine the objective function and constraint. The objective function is observation cost function, that can be written as:

$$f = n'C_1 + nC_2$$

The constraint is the variance of regression estimator in sample as follows:

$$g = S_y^2\left(\frac{1}{n} - r^2 \frac{n' - n}{nn'}\right) - \frac{\varepsilon^2}{Z_{\alpha/2}^2} = 0$$

Then we construct Lagrange function as in equation (5):

$$\begin{aligned} L &= n'C_1 + nC_2 + \lambda\left(S_y^2\left(\frac{1}{n} - r^2 \frac{n' - n}{nn'}\right) - \frac{\varepsilon^2}{Z_{\alpha/2}^2}\right) \\ &= n'C_1 + nC_2 + \lambda S_y^2 \frac{1}{n} - \lambda S_y^2 r^2 \frac{1}{n} + \lambda S_y^2 r^2 \frac{1}{n'} - \lambda\left(\frac{\varepsilon^2}{Z_{\alpha/2}^2}\right) \end{aligned} \tag{20}$$

Optimal condition for equation (20) is:

$$\frac{\partial L}{\partial n'} = C_1 - \lambda S_y^2 r^2 \frac{1}{n'^2} = 0 \tag{21}$$

$$\frac{\partial L}{\partial n} = C_2 - \lambda S_y^2 \frac{1}{n^2}\left(1 - r^2\right) = 0 \tag{22}$$

$$\frac{\partial L}{\partial \lambda} = S_y^2 \frac{1}{n} - S_y^2 r^2 \frac{1}{n} + S_y^2 r^2 \frac{1}{n'} - \frac{\varepsilon^2}{Z_{\alpha/2}^2} = 0 \tag{23}$$

From equation (21), we obtain:

$$\lambda = \frac{C_1 n'^2}{S_y^2 r^2} \tag{24}$$

From equation (22), we obtain:

$$\lambda = \frac{C_2 n^2}{S_y^2 (1-r^2)} \tag{25}$$

From equations (24) and (25), we obtain:

$$\frac{1}{n'} = \sqrt{\frac{C_1 (1-r^2)}{r^2 C_2 n^2}} \tag{26}$$

$$\frac{1}{n} = \sqrt{\frac{r^2 C_2}{C_1 n'^2 (1-r^2)}} \tag{27}$$

From equation (23), we get:

$$S_y^2 \frac{1}{n} - S_y^2 r^2 \frac{1}{n} + S_y^2 r^2 \frac{1}{n'} = \frac{\varepsilon^2}{Z_{\alpha/2}^2} \tag{28}$$

Then, substituting equation (27) to equation (28), we obtain:

$$\frac{S_y^2}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}} \left(r\sqrt{\frac{C_2}{C_1(1-r^2)}} - r^2 \sqrt{\frac{r^2 C_2}{C_1(1-r^2)}} + r^2 \right)$$

$$= n' \frac{S_y^2 r^2 + S_y^2 \sqrt{\left(\frac{C_2}{C_1}\right) r^2 (1-r^2)}}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}} = n' \tag{29}$$

By solving the above problem using Lagrange multipliers method, we get equation (29) which is a formula for calculating the number of optimal plots on first phase. We can write it as follows:

$$n'_{opt} = \frac{S_y^2 r^2 + S_y^2 \sqrt{\left(\frac{C_2}{C_1}\right) r^2 (1-r^2)}}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}} \tag{30}$$

Next, by substituting equation (26) to equation (28), we obtain:

$$\frac{S_y^2 (1-r^2) + S_y^2 \sqrt{\left(\frac{C_1}{C_2}\right) r^2 (1-r^2)}}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}} = n \tag{31}$$

Equation (31) is the formula for calculating the number of optimal plots on second phase. We can write it as follows:

$$n_{opt} = \frac{S_y^2 (1-r^2) + S_y^2 \sqrt{\left(\frac{C_1}{C_2}\right) r^2 (1-r^2)}}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}} \tag{32}$$

D. Implementation of Sampling Optimization by using Lagrange Multipliers Method

The method is implemented by using simulation of image interpretation data and field survey data. Image interpretation data is the data forest picture obtained from observations with remote sensing. The result of remote sensing is calculated by software until we get the diameter, density, and number of trees per plot, then we can also calculate tree volume per plot. Then, we check the result of image interpretations in the field.

To determine the potential of a forest, it is impossible to observe all objects in forest. Thus, we need to take some samples. In previous research of Fathia Amalia R. D, she takes 76 plot samples for first phase sampling which is in image interpretation and 38 plot samples for second phase without knowing whether the number of samples is optimum or not. In this paper, we need to calculate the optimal number of samples in image interpretation and in the field. Samples were observed in the form of plots where the plot consists of several trees.

Data of previous observation result can be used for calculating the optimal number of samples which must be observed on image interpretation and field. We use data from FMU Perum Perhutani Madiun II, East Java, which includes data from FSMU Dagangan and Dungus. For calculating the number of optimal samples, we use the following parameters:

TABLE I
SUM AND AVERAGE OF TREE VOLUME.

Parameter (m^3/0.1 ha) \ Location	FSMU Dagangan	FSMU Dungus
Sum of V_{image}(n samples) $\times V_{field}$	18864.1143	40627.7806
Sum of V_{image} (n samples)	831.99	1147.18
Sum of V_{field}	838.01	1131.31
Sum of V_{image}^2 (n samples)	18701.9755	40802.37
Sum of V_{field}^2	19103.4571	42398.9871
Average of V_{field}	22.05289	29.77132
Average V_{image}	22.60158	42.47013
Average V_{image} (n samples)	21.89447368	30.18894737

Observation cost consists of two types: image observation cost and field observation cost. Image observation cost is the total of cost which is used to buy image, image processing cost, and image map printing cost. Field observation cost is included in transportation cost, employee salary and etc. So that, we obtain the cost per hectare:

TABLE II
OBSERVATION COST.

Cost (Rp/ha) \ Location	FSMU Dagangan	FSMU Dungus
Image Interpretation	22.145	22.148
Field	363.158	363.157

Then to determine the optimal number of samples in the first phase (n'_{opt}) and second phase (n_{opt}), we calculate S_y^2 value first by using formula in equation (3), and also calculate r by using formula in equation (19). Next, we calculate n'_{opt} by using formula in equation (30) and calculate n_{opt} by using formula in equation (32).

We calculate all step by using Matlab, for location FSMU Dagangan we obtain optimal number of samples that must be observed for first phase n'_{opt} is 149 plots image interpretation and number of samples that must be observed for the second phase n_{opt} is 14 plots field survey. With the same way, for FSMU Dungus the number of optimal samples that must be observed is 153 plots image interpretation and 20 plots field survey.

IV. Conclusions

Based on analysis result and discussion, we obtain the following conclusions:

1) Result from formula analysis in sampling and optimization by using Lagrange multipliers method, we obtain the number of optimal samples in the formula for first phase (n') and second phase (n) is:

$$n'_{opt} = \frac{S_y^2 r^2 + S_y^2 \sqrt{\left(\frac{C_2}{C_1}\right) r^2 \left(1 - r^2\right)}}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}}$$

$$n_{opt} = \frac{S_y^2 \left(1 - r^2\right) + S_y^2 \sqrt{\left(\frac{C_1}{C_2}\right) r^2 \left(1 - r^2\right)}}{\frac{\varepsilon^2}{Z_{\alpha/2}^2}}$$

where,

- S_y^2 : variance(y) from the second phase sample (n)
- r : correlation coefficient
- C_1 : cost of first phase sampling
- C_2 : cost of second phase sampling
- ε : error in estimation
- $Z_{\alpha/2}$: value of random variables in standard normal distribution

2) From the calculation results, the number of optimal samples in image interpretation and field survey with FSMU Dagangan data, we obtain the number of optimal plots that must be observed in image interpretation is 149 plots and in field survey is 14 plots. The other side, with FSMU Dungus data we obtain the number of optimal plots that must be observed in image interpretation is 153 plots and in field survey is 20 plots. So that, if the number of samples is suitable with that calculation result, then we obtain an optimal sampling.

References

[1] R. Walpole, *Pengantar statistika, edisi ke-3 (Introduction to statistics)*. PT Gramedia Pustaka Utama, 1990.

[2] P. Malamassam, *Modul Mata Kuliah Inventarisasi Hutan*. Makassar: Hasanuddin University, 2009.

[3] K. Kitikidou, "Optimizing forest sampling by using Lagrange multipliers," *American Journal of Operations Research*, vol. 2, no. 1, pp. 94–99, 2012.

[4] S. Makridakis, S. Wheelwright, and V. McGee, *Metode dan Aplikasi Peramalan Jilid 1 (Ir. Untung Sus Ardiyanto, M. Sc. & Ir. Abdul Basith, M. Sc. Terjemahan)*. Jakarta: Erlangga, 1999.

[5] P. de Vries, *Sampling Theory for Forest Inventory*. Wageningen: Wageningen Agricultural University, 1986.

[6] D. Luknanto, *Pengantar Optimasi Non Linier*. Yogyakarta: Gajah Mada University, 2000.