CHAPTER 3

RESEARCH METHODOLOGY OF THE STUDY

==========

3.1 Primary Data Collection: Selection of the Area

The present study investigated the socio-economic status of Tharu Tribe with special reference to Lakhimpur Kheri district of Uttar Pradesh. The total population of Tharu Tribe in Uttar Pradesh is 105,291 and in case of Lakhimpur Kheri district it is 53,375 (Census of India, 2011), which is highest among all the districts in Uttar Pradesh. Tharu population is mainly concentrated to Palia and Nighasan block as the population in Palia is 40,468 and 8,000 in Nighasan. The overall literacy rate of Tharu Tribe is 56.9 percent in Kheri which is less than state average literacy of the district i.e. 60.6 percentage (Census, 2011).

3.2 Research Design

In the present study exploratory cum descriptive research design is adopted. This study is empirical in nature.

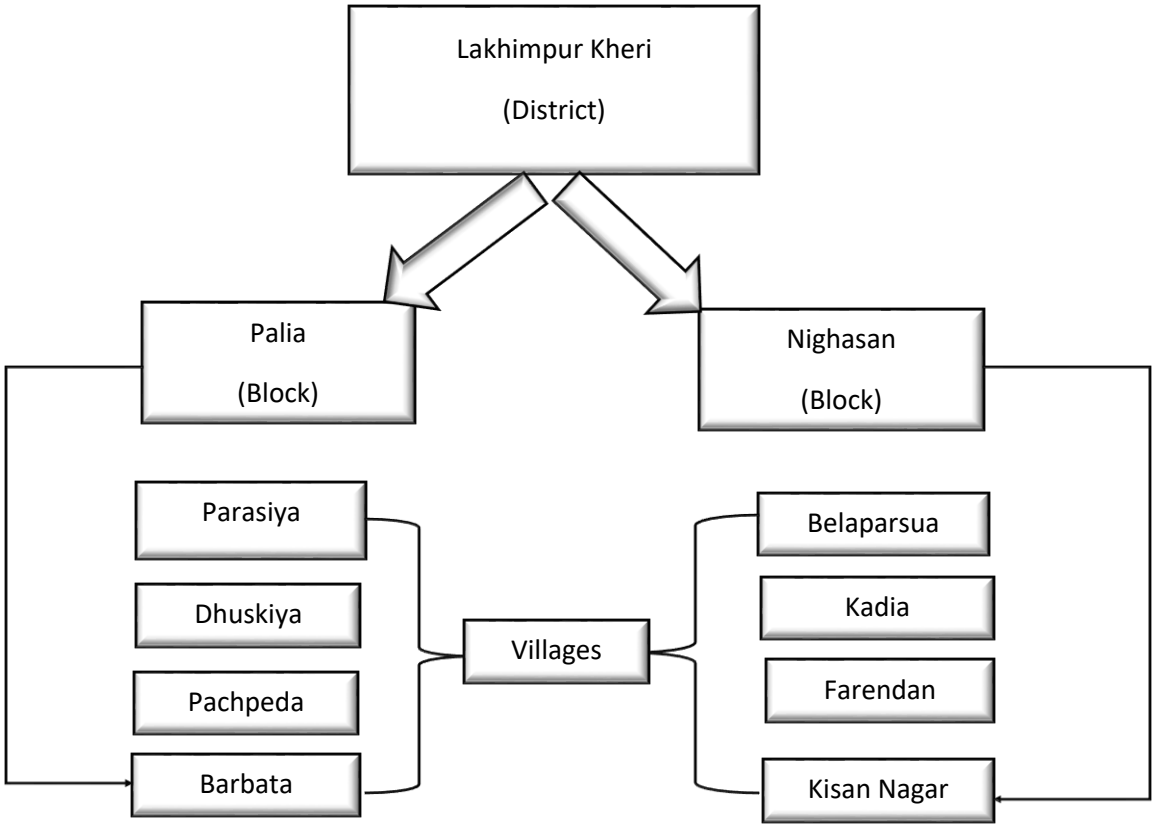
3.3 Sample and Sampling Design

For primary data collection, non-probability sampling viz. quota sampling has been followed. Lakhimpur Kheri district of Uttar Pradesh has been selected on the basis of highest population in terms of Tharu tribe. Further, two most populated tribal blocks have been selected and from this 248 household from Palia block and 62 household from Nighasan block has been selected by assigning quota. After the division of sample into blocks randomly four village from each block has been selected for study. The selected village were Parasiya, Barbata, Pachpeda and Dhuskia from Palia block and Belaparsua,

Kadiya, Kisannagar and Farendan from Nighasan block. The assigned quota has been fulfilled from the villages of the respective blocks. The selection of villages from blocks and then sample household from villages has been presented by flow chart.

Table 3.3.1	
Village	Number of Sample Household
Parasiya	103
Pachpeda	51
Dhuskiya	50
Barbata	44
Belaparsua	42
Kadiya	3
Kisannagar	8
Farendan	9

3.3.2 Flow-chart of the Sample Area



3.4 Data and Data Sources

The study is based on both primary as well as secondary data. The primary data have been collected through field survey by using well prepared household schedule. The secondary data for the study has been collected from various government sources such as Census Abstract, Statistical Profile of STs (2013) and Socio-economic Caste Census, 2011.

3.5 Area and Period of Study

The area under the study is Lakhimpur Kheri District of Uttar Pradesh. More specifically the samples are taken from two block namely Palia and Nighasan. The data pertain to the reference period of March to April 2019.

3.6 Data Collection

For data collection, well-prepared schedule has been used. The schedule is designed in such a manner to cover the social, economic and demographic variables. In addition to that, there are several questions in the schedule which are helpful in assessing their access to government school, hospital and several welfare schemes (Appendix I). The primary data has been collected by direct interview of respondent using the study schedule. The information collected with the interview tried to explore the social status, occupation, land ownership, educational status, health status, status of SHGs, access to bank, and access to various government scheme for toilet, electricity, house, pension, schools, and hospitals.

3.7 Data Analysis

3.7.1 Scenario of Socio-economic Condition of Tribe: The scenario of socio-economic status of Tharu Tribe has been analyzed by using frequency and percentage method. Here,

socio-economic status has been assessed by housing condition, land ownership, own vehicles, monthly household income, access to bank, school, hospital and various government schemes.

3.7.1.1 Frequency and Percentage Method: The socio-economic status of Tharu tribe has been explained by calculating frequency and percentage. Frequency, is the rate at which something occurs over a particular period of time or in a given sample.

Percentage method is used for obtaining different figures from the absolute value of households for different variable.

$$S.P. = \frac{n^{th}}{N} \times 100$$

where, S.P. = simple percentage

N = total number of observations, n= each group of the observation

3.7.2 Socio-economic Factor Influencing the Household Income

There is a greater influence of socio-economic factors on household income (Tuyen, 2015). The explanatory variables incorporated in the model include socio-economic factors based on the existing literature work and field knowledge of the researcher. The explanatory variable taken in the study are family size, gender of head of household, occupation, land ownership, educational level of household and SHGs membership.

Econometric Model: This model estimated the household income for a cross-sectional sample of 310 household. We have used multiple regression model to analyze the effect of the five explanatory variables to the target variable, and out of these five, two variables such as education and occupation have used 4 dummies for each.

The formula for multiple regression is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots\ldots\ldots\ldots\ldots + \beta_k X_{ki} + \varepsilon i \quad \ldots.(1)$$

where,

Y_i = ith observation of the dependent variable Y, i= 1, 2……n

X_j = independent variables, j= 1, 2 ……..k

X_{ji} = i^{th} observation of the j^{th} independent variables

β_0 = intercept term

β_j = slope coefficient for each of the intendent variable

$\mathcal{E}_i$ = error term for the i^{th} observation

n = number of observation

k = number of independent variables

In present study eq^n…

$$lnY_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 D_{1i} + \beta_4 D_{2i} + \beta_5 D_{3i} + \beta_6 D_{4i} + \beta_7 D_{5i} + \beta_8 D_{6i} + \beta_9 D_{7i} + \beta_{10} D_{8i} + \mathcal{E}i$$

here, Yi = Household total income

X_1 = Family Size (Number of members in family)

X_2 = Land [Land ownership of household (in bigha)]

D_1= 1, if person is male otherwise 0

D_2= 1, if person is government servent oterwise 0

D_3= 1, if person is traditional farmer oterwise 0

D_4=1, if person is subsistant farmer oterwise 0

D_5= 1, if household have primary education otherwise 0

D_6= 1, if household have secondary education otherwise 0

D_7=1, if household have higher education otherwise 0

D_8= 1, if household have SHGs membership otherwise 0

$\mathcal{E}i$ = error term for the ith observation

Correlation formula is:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{cov(X,Y)}{StandardDev(X) \times StandardDev(Y)}$$

Correlation matrix is:

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ r_{31} & r_{32} & 1 & \cdots & r_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{pmatrix}$$

$$\text{where, } r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(X_{ij}-\overline{X_j})(X_{ik}-\overline{X_k})}{\sqrt{\sum_{i=1}^{n}(X_{Ij}-\overline{X_j})^2}\sqrt{\sum_{i=1}^{n}(X_{ik}-\overline{X_k})^2}} = \begin{cases} 1 \ if \ \ j = k \\ r_{jk} if \ \ j \neq k \end{cases}$$

To analyze the degree of multicollinearity the size of VIF has been analyzed. It is related to the multiple correlation coefficient R_i between variable j and the other variables.

Variance Inflation Factor (VIFs) for $\widehat{\beta_\iota}$

$$VIFs\ (\widehat{\beta_\iota}) = \frac{1}{(1-R_i^2)}$$

where, R_i^2 is the unadjusted R^2, i=1,2,.....k

the rule of thumb is if VIF>5 then multicollinearity is an issue.

Heteroscedasticity Formula: One of the characteristics made about errors term in simple ordinary least square regression is that the residual has the same but unknown variance. This is known as constant variance or homoscedasticity. When this assumption is violated, the problem is known as heteroscedasticity. In this study we have used Breusch Pagan Test was introduced by Trevor Breusch and Adrian Pagan in 1979. In this model they assumed that the error terms are normally distributed. It tests whether the variance of the errors from a regression is dependent on the values of the independent variables.

$$H_0: Var\left(\frac{u}{x_1, x_2, x_3 \ldots\ldots\ldots\ldots x_k}\right) = Var\left(\frac{u}{X}\right) = \sigma^2$$

$$Var\left(\frac{u}{X}\right) = E\left(\frac{u^2}{X}\right) - \left[E\left(\frac{u}{X}\right)\right]^2 = E(\frac{u^2}{X})$$

$$\Longrightarrow \qquad E\left(\frac{u^2}{x_1, x_2, \ldots\ldots\ldots\ldots x_k}\right) = E(u^2) = \sigma^2$$

where, u^2 is a squared of error term