Chapter–5

Research Methodology: The Approach and Tools

A Biswas, Moumita Gupta, G C Mishra and S K Acharya

The discussion on the methodology has been made to understand the concepts, methods and techniques, which are utilized to design the study, collect the information, analyze the data and interpret the findings for revelation of truth and formulation of theories. This chapter deals with the research methodology that had been adopted for the purpose of the present study. However, the entire discussion has been made under the following sub-themes:

5.1 Locale of study

5.2 Sampling design

5.3 Pilot study

5.4 Variables and measurements

5.5 Method of data collection

5.6 Statistical tools used for analysis and interpretation of data.

5.1 Locale of the study (research location):

Keeping in view the paucity of time, resources and current socio-political situation the present study was conducted at Hezamara block of West district (Tripura). The district, block and village were selected purposively due to the following reasons:

a. The research's close familiarity with respect to ar ea, people, their culture and local dialects.

b. The concerned areas are easily accessible to the researcher in term of transportation and place of residence.

c. The opportunity of getting good response from the respondents

d. The study will help the researcher in future for conducting various extension programme and activities if he serves the agricultural department of state in future.

5.2 Sampling design

The purposive as well as simple random sampling techniques were adopted for the present study. It may be termed as multistage random sampling procedure. The districts, blocks and villages were purposively selected for the study. The West district and the block Hezamara were considered. Under the Hezamara block Sharat chowdhury para village was selected. From Sharat chowdhury para village 96 samples have been selected out of 1500 bamboo grower.

Table 5.1: The Sampling Paradigm and Approaches

Level	***Locale***	***Name***	***Approach***	***Selection criteria***	
1.	**STATE**	**TRIPURA**	**PURPOSIVE**		1. Booming bamboo trade.
					2. Bamboo is an emerging economy for Tripura.
					3. Bamboo generates income and livelihood for
					rural people.
					4. The research would provide an empirical
					support to the management of bamboo
					economy.

2.	**DISTRICT**	**WEST**	**PURPOSIVE**	1.	Offers better accessibility.
				2.	The intensity of bamboo cultivation is one of the highest.
				3.	The district provides unique bamboo based economy.
3.	**BLOCK**	**HEZAMARA**	**PURPOSIVE**	1.	Offers better accessibility.
				2.	The intensity of bamboo cultivation is one of the highest.
				3.	The block provides unique bamboo based economy.
4.	**VILLAGE**	**SHARAT CHOWDHURY PARA**	**PURPOSIVE**	1.	Offers better accessibility.
				2.	The intensity of bamboo cultivation is one of the highest.
				3.	The village provides unique bamboo based economy.

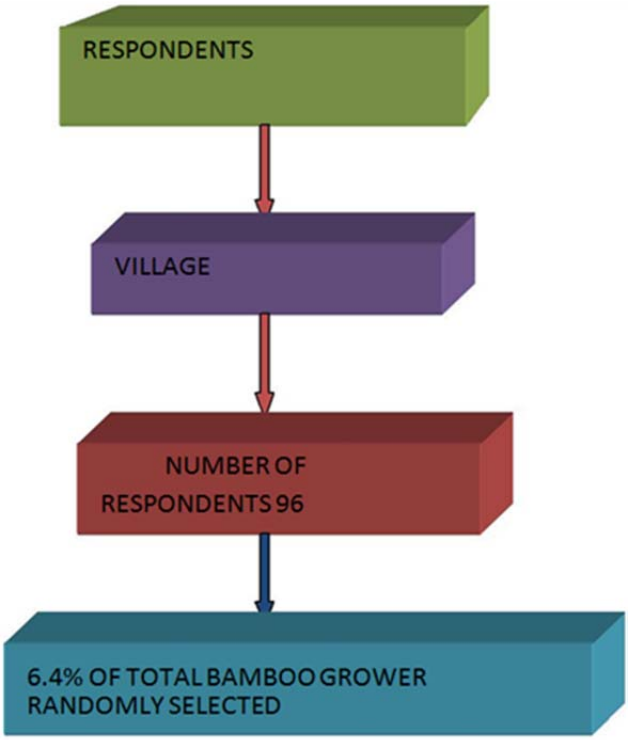


Figure No. 5.1: Selection Paradigm of Respondents from Village sharat Chowdhury Para

5.3 Pilot study

A pilot study was conducted in the selected villages before constructing the data collecting devices. In course of this survey informal discussion was carried out with some bamboo growers and other resource persons of the localities. An outline of socio-economic background of the farmers of the concerned village, their opinion towards different types of bamboo plantation, their productivity, and livelihood generated from bamboo and other agricultural enterprise, income generation were obtained that helped in the construction of reformative working tools.

5.4 Variables and their measurement

Several researchers pointed out that the behaviour of an individual was understood more in depth if one has the knowledge of some variables, which comprised the constructed world of reality within which an individual received the stimuli and acts. The **socio-personal, agro-economic, and Techno-managerial** variables are such type of variables, which determine the behaviour of an individual. Appropriate operationalisation and measurement of the variables help the researcher to land upon the accurate conclusion. Therefore, the selected variables for this study had been operationalised and measured in following manner:

Independent variables Socio-personal variables

Age (X_1)

It denotes the chronological age, years and months elapsed since birth of the respondent. It was measured through counting the chronological age.

Education (X_2)

Education is the factor that has been conceived in terms of acquisition of knowledge and skill formality in school. The values ascribed the class in which he or she studied or studying.

Family size (X_3)

It denotes the total number of persons living in the respondents' house under a single household.

Material possession (X_{10})

The respondents were asked whether they possessed some of the household materials like cycle, pakka house, motor bike, freeze, fan etc. The score assigned as in the case mentioned in the interview scheduled.

Training received(X_{15})

The respondents were asked how many days they received training regarding the technical knowhow of bamboo cultivation and also post harvest processing, value addition for making crafts etc. It is measured by adding total number of days they received training in each time.

Energy consumption(X_{16})

The respondents were asked how much amount of money they spend in energy consumption purpose. It is measured by the total amount of money they spend in electricity, L.P.G, firewood, kerosene, petrol.

Mass media exposure (X_{12})

It refers to the use of information source(s) both from outside and within the community for getting agricultural information. The score assigned as in the case mentioned in the interview scheduled.

Agro-economic variables

Homestead land(X_6)

Mass of land congruent to dwell and generally used for household purpose. It is measured in term of kani (1 ha= 6.25 kani) and divided by family size.

Land under agricultural crop(X_7)

The mass of land under field crops, measured in term of kani, which has been divided by family size.

Cropping intensity(X_8)

It refers to the proportion of acreage annually put under different crops to the total cropped area as expressed in percentage. The cropping intensity is calculated by

$$\text{Cropping intensity (\%)} = \frac{\text{Gross Cropped Area}}{\text{Net Cropped Area}} \times 100$$

Annual income before bamboo(X_{11})

The respondents were asked what annual income before bamboo cultivation was there. The annual income is measured by dividing the whole income by family size.

Techno-Managerial Variables

Average Cost of farm implements when purchased (X_4)

It is measured by the total cost of farm implements when purchased, divided by number of implements.

Average Cost of farm implements at present (X_5)

It is measured by the present total cost of farm implements, divided by number of implements.

Area under bamboo(X_9)

The mass of land area under bamboo cultivation measured in term of kani which has divided by family size.

Number of Rhizome planted(X_{13})

The respondents were asked how many number of rhizome of bamboo plant as seed material they planted in their planted.

Number of Rhizome grown to the fullest(X_{14})

The respondents were asked how many number of rhizomes grew up to full plant. It was measured by deducting the number of rhizome died due disease, low quality from the total number of rhizome planted.

Cost incurred in bamboo cultivation(X_{18})

The respondents were asked total how much money they have invested starting from fencing, weeding, and cultural practices in 3 years before selling. It is measured by adding total cost incurred in 3 year in cultural practices in bamboo cultivation.

Distance to market(X_{17})

The respondents were asked how far the market where they sell their produce is.

Mode of selling(X_{19})

The respondents were asked to whom they sell their produce**.** The score assigned as in the case mentioned in the interview scheduled.

Dependent variables

The appropriate operationalisation and measurement of the predicted variables help in concluding the study in a proper manner. This is a very interesting area of work in measuring the variables after conceptualizing them.

Family income from Bamboo enterprise (Y_1)

The family income from bamboo enterprise by the respondents. It is measured as-

$$\text{Family income from bamboo enterprise} = \frac{\text{Total income from bamboo enterprise}}{\text{Number of family member}}$$

Annual income from Agricultural enterprise (Y_3)

The family income from agricultural enterprise by the respondents, it is measured as-

$$\text{Family income from Agricultural enterprise} = \frac{\text{Total income from agricultural enterprise}}{\text{Number of family member}}$$

Productivity of Bamboo (Y_3)

Productivity of bamboo is measured as follows

$$\frac{\text{Total number of bamboo pole produced per year}}{\text{Family size}}$$

Livelihood generated from Bamboo enterprise (Y_4)

Livelihood generated from bamboo enterprise is measured by per year. Wages and Mandays generated from bamboo enterprise.

5.5 Methods of data collection

The primary data in the present study were collected directly from the farmers with the help of structured schedule through personal interview methods. Only the functional head of the household were taken as respondents for the study.

The personal interview method was followed during the month of May and October 2013 to collect the relevant information from targeted respondents.

5.6 Statistical tools used for analysis and interpretation of data.

After collection of data, data were processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. Processing implies editing, coding, classification, and tabulation of collected data. The main Statistical techniques and tool used in the present study-

Mean

The mean is the arithmetic average and is the result obtained when the sum of the value of individual in the data is divided by the number of individuals in the data. Mean is the simplest and relatively stable measure of central tendency. The mean reflects and is affected by every score in the distribution. Thus, extreme scores affect the mean.

Calculation of mean from grouped data

When the data are expressed in a frequency distribution (grouped), the mean calculated by the formula**.**

$$\bar{x} = \frac{\sum fx}{N}$$

Where,

$\bar{\bar{x}}$ mean of the distribution

f = frequency of the class

x = class value of midpoint of the class internal

N = number of observations

Standard deviation

Standard deviation is the square root of the arithmetic mean of the squares of the deviations, the deviations being measured from the arithmetic mean of distribution. It is commonly denoted by the symbol sigma. It is less affected by sampling errors and is more stables measure of dispersion. The Standard deviation of the data grouped in the form of frequency distribution is computed by the formula-

$$a=\sqrt{\frac{\sum f d^2}{N}}$$

Where,

I = frequency of the class

d = deviation of the mid-value of the class from the population mean N = total number of observations.

Coefficient of variation

A measure of variation which is independent of the unit of measurement is provided by the Coefficient of variation. Being unit free, this is useful for comparison of variability between different populations. The Coefficient of variation is standard deviation expressed as per centage of the mean and is measured by the formula-

$$CV=\frac{Standard\ deviation\ (\sigma) x\ 100}{Mean}$$

Correlation

When an increase or decrease in one variety is accompanied by an increase or decrease in the other variety, the two are said to be correlated and the phenomenon is known as correlation. Correlation coefficient (r) is a measure of the relationship between two variables, which are at the interval or ratio level of measurement and are linearly related. A Person product-moment 'r' is computed by the formula.

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum x^2 - (\sum x^2)][N\sum Y^2 - (\sum Y)^2]}}$$

Where,

X and Y = original scores in variables X and Y

N = number of paired scores

∑XY = each X multiplied by its corresponding Y, then summed

∑X = sum of X scores

$\sum X^2$ = each of X squared, then summed

$(\sum X)^2$ = sum of X scores squared

∑Y = sum Y scores

$\sum Y^2$ = each of Y squared, then summed

Regression

The correlation coefficient only expresses association and by itself tells us nothing about the causal relationships of the variables. Thus, purely from the knowledge that two variables x and y are correlated, we cannot say whether variation in x is the cause or the results from mutual dependence of the two varieties or form common causes affecting both of them. Similarly, the mere existence of a high value of correlation coefficient is not necessarily of an underlying relationship between the two variables.

The underlying relation between y and x in a bi variant population can be expressed in the form of a mathematical equation known as regression equation and is said to represent the regression of the variety y on the variety x.

If Y is the dependent variable and X is the independent variable, then the linear regression equation can be written as -

$$Y = a + bx$$

The values of a and b can be obtained by the method of least squares which consists of minimizing the expression.

$(\sum y_i - a - bx_i)^2$ with respect to a and b. The value of a and b are

$$b = \frac{\sum XY - \frac{(\sum x)(\sum y)}{n}}{\sum X^2 - \frac{(\sum x)^2}{n}}$$

$$a = \bar{y} - b\bar{x}$$

The regression line can now be written as

$$Y = \bar{y} + b\bar{x} + bx \text{ or } y - \bar{y} = b(x - \bar{x})$$

Where, b is the regression coefficient.

Path analysis

The terms 'path analysis' was first introduced by t he biologist Sewall Wright in 1934 in connection with decomposing the total correlation between two variables in the casual system. The technique of path analysis is based on a series of multiple regression analysis with the added assumption of causal relationship between independent and dependent variables. This technique lays relatively heavier emphasis on the heuristic

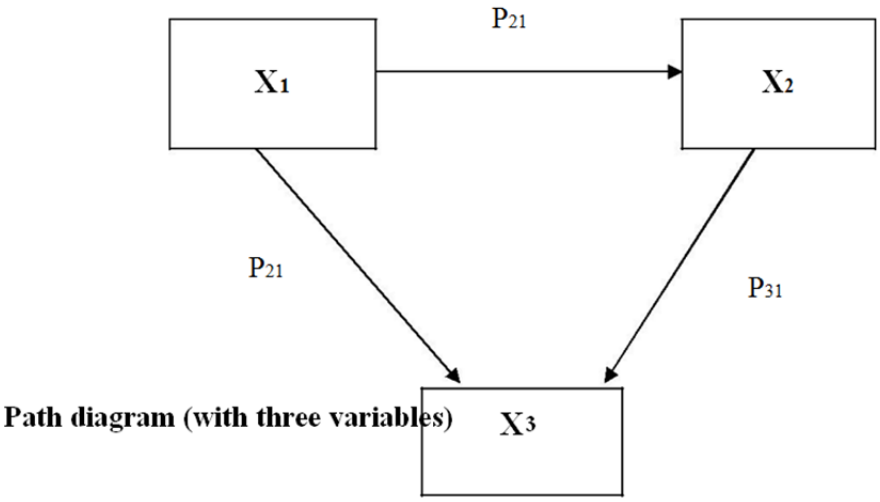
use of visual diagram, technically described as a path diagram. An illustrative path diagram showing interrelationship between father's education, father's occupation, son's education, son's first a nd son's first and son's present occupation can be shown.

Path analysis makes use of standardized partial regression coefficients (known as beta weights) as effect coefficients. In linear additive effects are assumed, then through path analysis a simple set of equations can be build up showing how each variable depends on preceding variables. The main principle of path analysis is that any correlation coefficient between two variables, or a gross or overall measure of empirical relationship can be decomposed into a series a paths: separate path of influence leading through chronologically intermediate variable to which both the correlated variables have links.

The merit of path analysis in comparison to correlation analysis is that it makes possible the assessment of the relative influence of each antecedent of explanatory variable on the consequent or correlation variable by first making explicit assumptions underlying the causal connections and then by elucidating the indirect effect of the explanatory variables.

The use of the path analysis technique requires the assumption that there are linear additives, a symmetry relationship among a set of variables which can be measured at least on a quest interval scale. Each dependent variable is regarded as determined by the variable preceding it in the path diagram, and a residual variable defined as uncorrelated with other variables, is postulated to account for the unexplained portion of the variance in the dependent variable. The determining variables are summed for the analysis to be given (exogenous in the model).

We may illustrate the path analysis technique in connection with a simple problem of testing a causal model with three explicit variables as shown in the following path diagram:



Path diagram (with three variables)

The structural equation from the previous can be written as:

$X_1 = e_1$

$X_2 = P_{21}X_1 + e_2 = px + e$

$X_3 = P_{53}X_2 + P_{32}X_2 + e_3$

X1 and X variable are measured as deviation from their respective means. P_{21} may be estimated from the simple regression of X_2 on X_1, i.e., $b_{21}X_1$ and P_{31}may be estimated from the regression of X on X_2 and X_1 as under:

$X = P_{31}X_1 + b_{21}x_2$

Where, $b_{21}x_2$ means the standardized partial regression coefficient for predicting variable 1 when the effect of variable 2 is held constant.

In path analysis the beta co-efficient indicates the direct of X1 (j=1, 2, 3,..p) on the dependent variable. Squaring the direct effect yields the proportion of variance on the dependent variable Y which is due each of the number of independent variable X_1 (j=1,2,3...p). After calculating the direct effect one may obtain a summary measure of the total indirect of X_1 on the dependent

variable Y by subtracting from the correlation coefficient r_{yxj} the beta co-efficient b i.e.

Indirect effect X_1 on $y=Cjy=r_{yxj}-b_1$

For all j=1, 2, 3...p

Factor analysis

Many statistical methods are used to study the relation between independent and dependent variables. Factor analysis is different; it is used to study the patterns of relationship among many dependent variables, with the goal of discovering something about the nature of the independent variables that affect them, even though those independent variables were not measured directly. Thus answers obtained by factor analysis are necessarily more hypothetical and tentative than is true when independent variables are observed directly. The inferred independent variables are called factors.

Concepts used in factor analysis

Some important concepts used in factor analysis are explained by Kothari (1996) as follows-Factor-A factor is an underlying dimension that accounts for several observed variables. Factor is a hypothetical construct or classification. There may be one or more factors, depending upon the nature of the study and the number of variables involved in it.

Factor loading - Factor loadings are those values which explain how closely the variables are related to each one of the factor discovered. Factor loadings work as key to understanding what the factor mean. It is the absolute size (rather the sign, plus or minus) of the loading that is important in the interpretation of a factor.

Since the factors happen to be linear combinations of data, the co-ordinates of each observation or variables is measured to obtain what are called factor loading. Such factor loading represent the correlation between the particular variable and the factor, and are usually placed in a matrix of correlation between the variables and the factor.

Canonical correlation analysis

Multiple regression analysis is a multivariate technique which can predict the value of a single (metric) dependent variable from a linear function of a set of independent variables. For some research problems, however, interest may not center on a single dependent variable; rather, the researcher may be interested in relationships between sets of multiple dependent and multiple independent variables. **Canonical correlation analysis** is a multivariate statistical model that facilitates the study of interrelationships among sets of multiple dependent variables and multiple independent variables. Whereas multiple regression predicts a single dependent variable from a set of multiple independent variables, canonical correlation simultaneously predicts multiple dependent variables from multiple independent variables. Canonical correlation places the fewest restrictions on the types of data on which it operates. Because the other techniques impose more rigid restrictions, it is generally believed that the information obtained from them is of higher quality and may be presented in a more interpretable manner. For this reason, many researchers view canonical correlation as a last-ditch effort, to be used when all other higher-level techniques have been exhausted. But in situations with multiple dependent and independent variables, canonical correlation is the most appropriate and powerful multivariate technique. It has gained acceptance in many fields and represents a useful tool for multivariate analysis, particularly as interest has spread to considering multiple dependent variables.

Analyzing Relationships with Canonical Correlation

Canonical correlation analysis is the most generalized member of the family of multivariate statistical techniques. It is directly related to several dependence methods. Similar to regression, canonical correlation's goal is to quantify the strength of the relationship, in this case between the two sets of variables (independent and dependent). It corresponds to factor analysis in the creation of composites of variables. It also resembles discriminant analysis in its ability to determine independent dimensions (similar to discriminant functions) for each variable set, in this situation with the objective of producing the maximum correlation between the dimensions. Thus, canonical correlation identifies the optimum structure or dimensionality of each variable set that maximizes the relationship between

independent and dependent variable sets. Canonical correlation analysis deals with the association between composites of sets of multiple dependent and independent variables. In doing so, it develops a number of independent **canonical functions** that maximize the correlation between the **linear composites**, also known as **canonical variates**, which are sets of dependent and independent variables. Each canonical function is actually based on the correlation between two canonical variates, one variate for the dependent variables and one for the independent variables. Another unique feature of canonical correlation is that the variates are derived to maximize their correlation. Moreover, canonical correlation does not stop with the derivation of a single relationship between the sets of variables. Instead, a number of canonical functions (pairs of canonical variates) may be derived. The following discussion of canonical correlation analysis is organized around a six-stage model-building process. The steps in this process include (1) specifying the objectives of canonical correlation, (2) developing the analysis plan, (3) assessing the assumptions underlying canonical correlation, (4) estimating the canonical model and assessing overall model fit, (5) interpreting the canonical variates, and (6) validating the model.