Chapter–6

Research Methodology

K.P. Kaleon 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, collects the information, analyze the data and interpret the findings for revelation of truth and formulation of theories. This chapter deals with the research methodology, had been adopted for the purpose of the present study.

However, the entire discussion has been made under the following sub-themes

- Locale of research
- Sampling design
- Pilot study
- Variables and their measurements
- Method of data collection
- Statistical tools used for analysis of data

1. LOCALE OF RESEARCH

The present study was conducted at Phodong forest block (Mangan sub-division) of North District (Sikkim). The districts block and villages were selected purposively due to the following reasons:

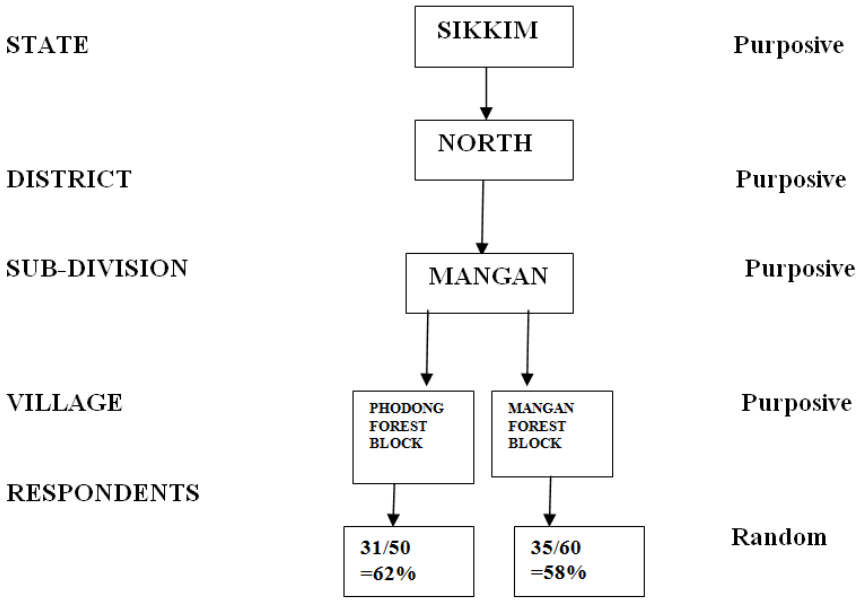
1. The characters and the factor under study have been well discernible in this area
2. The researcher's close familiarity with respect to area, people officials and local dialects;
3. The ample opportunity to generate relevant data due to the close proximity of the area with the research and extension wing of the State Agricultural;
4. The highly cooperative, responsive respondents;
5. The profuse scope to get relevant information regarding adoptionrejection and discontinuance of agricultural technology;
6. Experienced, well versed, venturesome and risk bearing farm entrepreneurs;
7. Easy accessibility of the area;
8. The study would help the researcher to conduct diversified extension programmes and activities in future.

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, sub-division and villages were purposively selected for the study. The North district and the sub-division Mangan were considered. Under the Mangan Sub-division Phodong forest block village was selected. An exhaustive list of respondents was prepared with the help of block officials for villages. From the prepared list to hundred respondents were selected randomly from each village for the final data collection.

Sampling Scheme (Multistage Random Sampling)

ISBN: 978-81-930585-3-4



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 farmers, local leaders and extension agents of the localities. An outline of socio-economic

background of the farmers of the concerned villages, their opinion towards different types of technology socialization process, innovation-decision process, adoption, non adoption, discontinuance and rejection were obtained that helped in the construction of reformative working tools.

4. VARIABLES AND THEIR MEASUREMENTS

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, socio-psychological and communication 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

1. Age(X1)

In all societies, age is one of the most important determinants of social status and social role of the individual. In the present study, the number of years rounded in the nearest whole number the responded lived since birth at the time of interview, was taken as a measure of age of the farmer.

2. Education(X2)

Education may be operationalised as the amount of formal schooling attained/ literacy acquired by the responded at the time of interview. Education is instrumental in building personality structure and helps in charging one's behaviour in social life.

3. Family size (X3)

Family size was operationalised as the number of members in the individual farmer's family.

ISBN: 978-81-930585-3-4

4. Media Interaction (X4)

The frequency of interaction and exposure to different media comprising of mass, local and interpersonal that has been measured through multiplying the frequency with no. off media. The digital value was considered as the variable value.

5. Per Capita Holding Size (X5)

It is the total size of the landmass under family ownership where in the respondent is the head of the family.

6. Cropping Intensity (X6)

Cropping intensity has been operationalised as the proportion of total annual cropped area to the size of holding expressed in percentage. The cropping intensity was calculated by the formula:

$$\text{Cropping intensity} = \frac{\text{Total annual cropped area in bigha}}{\text{Size of holding in bigha}} \times 100\%$$

7. Technology Socialization Status (X7)

It has been measured by the summation of percentiles consisting of adoption, rejection, discontinuance, and reinvention over some selected technology.

8. Family Income (X8)

It is the income from three basic sources agriculture, animal and service sectors divided by total Family size**.**

9. Expenditure after Health (X9)

It is the average expenditure (total expenditure/family size) incurred by a family on health.

10. Animal Health Mentoring (X10)

It is the average expenditure (total expenditure/family size) incurred by a family on Animal health mentoring.

11. Location of the Market (X11)

It is the working distance to a set of strategic points and an average market of all the distances has been calculated as the digital value for the variable.

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.

In the present study, the study gave insight into the contemplation part of the farmer's psyche, which was dealing with the agriculture innovation. It considered the all post-adoption phenomena under a single continuous contemplating process. For this reason the measurement of these variables had carried out in following manner.

1. Climate Change Perception (Y1)
2. Yield Change Perception (Y2)
3. Water Bodies Perception (Y3)
4. Health Problem Perception (Y4)
5. Species Decline Perception (Y5)
6. Perception Indicator Change (Y6)
7. Landslide Perception (Y7)
8. Distance Perception (Y8)
9. Comprehensive Climate Change Perception

All the above variables were measured using structured interview schedule and questionnaire method.

ISBN: 978-81-930585-3-4

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 heads of the household were taken as respondents for the study.

The personal interview method was followed during the month of December, 2011 to June 2012 to collect the relevant information from targeted respondents.

6. STATISTICAL ANALYSIS AND INTERPRETATION OF DATA

(ANALYTICAL TOOLS)

The role of statistics in research is to function as a tool in designing research, analyzing its data and drawing conclusions of there form. Most research studies result in a large volume of raw data, which must be suitably reduced so that the same can be read easily and can be used for further analysis. Clearly the science of statistics cannot be ignored by any research worker, even though he may not have occasion to use statistical method in all their details and ramifications.

After collection of data, data were processed and analysed 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 tools used in the present study were-

Mean, Standard deviation, Coefficient of variation, Coefficient of correlation, Regression, Multiple regression (Step-wise regression and Backward regression), Path analysis, Factor analysis, Principal component Analysis and Canonical constant analysis.

- **Mean**

Measure of central tendency (or statistical averages) tells us the point about which items have a tendency to cluster. Such a measure is considered as the most representative figure for the entire mass of data. Measure of central tendency is also known as statistical average. Mean, median and mode are the most popular averages. Mean, also known as arithmetic average, is the most common measure of central tendency and may be defined as the value, which we get by dividing the total of the values of various given items in a series by the total number of items. We can work it out as under:

$$\text{Mean or } (X) = \sum \frac{Xi}{n}$$

Where, X = The symbol we use for mean (pronounced as X bar)

Σ = Symbol for summation

Xi = Value of the ith item X, i = 1, 2, …….., n

N = total number of items

Mean is the simplest measurement of central tendency and is a widely used measure. Its chief use consists in summarizing the essential features of a series and in enabling data to be compared. It is amenable to algebraic treatment and is used in further statistical calculations. It is a relatively stable measure of central tendency. But it suffers from some limitations viz., it is unduly affected by extreme; it may not coincide with the actual value of an item in a series, and it may lead to strong impressions, particularly when the item values are not given with the average. However, mean is better than other averages, especially in economic and social studies where direct quantitative measurements are possible.

- **Weighted Mean**

The **weighted mean** is similar to an arithmetic mean (the most common type of average), where instead of each of the data points contributing equally to the final average, some data points contribute more than others.

ISBN: 978-81-930585-3-4

The notion of weighted mean plays a role in descriptive statistics and also occurs in a more general form in several other areas of mathematics.

If all the weights are equal, then the weighted mean is the same as the arithmetic mean. While weighted means generally behave in a similar fashion to arithmetic means, they do have a few counterintuitive properties, as captured for instance in Simpson's paradox.

The term **weighted average** usually refers to a weighted arithmetic mean, but weighted versions of other means can also be calculated, such as the weighted geometric mean and the weighted harmonic mean.

- **Mathematical definition**

Formally, the weighted mean of a non-empty set of data

$$\{x_1, x_2, \ldots, x_n\},$$

with non-negative weights

$$\{w_1, w_2, \ldots, w_n\},$$

is the quantity

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},$$

which means:

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}.$$

Therefore data elements with a high weight contribute more to the weighted mean than do elements with a low weight. The weights cannot be negative. Some may be zero, but not all of them (since division by zero is not allowed).

The formulas are simplified when the weights are normalized such that they sum up to 1, i.e. $\sum_{i=1}^{n} w_i = 1$.

For such normalized weights the weighted mean is simply $\bar{x} = \sum_{i=1}^{n} w_i x_i$.

Note that one can always normalize the weights by making the following transformation on the weights $w_i' = \frac{w_i}{\sum_{i=1}^{n} w_i}$. Using the normalized weight yields the same results as when using the original weights. Indeed,

$$\bar{x} = \sum_{i=1}^{n} w_i' x_i = \sum_{i=1}^{n} \frac{w_i}{\sum_{i=1}^{n} w_i} x_i = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

The common mean $\frac{1}{n}\sum_{i=1}^{n} x_i$ is a special case of the weighted mean where all data have equal weights, $w_i = w$. When the weights are normalized then $w_i' = \frac{1}{n}$.

ISBN: 978-81-930585-3-4

- **Standard deviation**

Standard deviation is the most widely used measure of desperation of a series and is commonly denoted by the symbol 'σ' (pronounced as sigma). Standard deviation is defined as the square root of the average of squares of deviations, when such deviations for the values of individual items in a series are obtained from the arithmetic average. It is worked out as under.

Standard deviation (σ) =$n\ X\ X\ i\ \Sigma(-)2$

- **Coefficient of variance**

When we divide the standard deviation by the arithmetic average of the series, the resulting quantity is known as coefficient of standard deviation, which happens to be relative measure, and is often used for comparing with similar measure of other series. When this coefficient of standard deviation is multiplied by 100, the resulting figure is known as coefficient of variation. Sometimes, we work out the square of standard deviation, known as variance, which is frequently used in the context of analysis of variation.

The standard deviation (along with several related measures like variance, coefficient of variation etc.) is used mostly in research studies and is regarded as a very satisfactory measure of dispersion in a series. It is amenable to mathematical manipulation because the algebraic signs are not ignored in its calculation (as we ignore in case of mean deviation). It is less affected by fluctuations of sampling. These advantages made standard deviation and its coefficient a very popular measure of the scatteredness of a series. It is popularly used in the context of estimation and testing of hypotheses.

- **Coefficient of correlation**

So far we have dealt with those statistical measures that we use in context of univariate population i.e., the population consisting of measure of only one variable. In case of bivariate or multivariate populations, we often wish to know the relation of the two and/or more variables in the data to one another.

Karl Pearson's coefficient of correlation (or simple correlation) is the most widely used method of measuring the degree of relationship between two variables. This coefficient assumes the following:

(i) That there is linear relationship between the two variables;

(ii) That the two variables are causally related which means that one of the variable is independent and the other one is dependent; and

(iii) A large number of independent causes are operating in both variables so as to produce a normal distribution.

Karl Pearson's coefficient of correlation

Where, Xi = ith value of X variable

X = mean of X

Yi = ith value of Y variable

Y = mean of Y

n = number of pairs of observations of X and Y

σx = Standard deviation of X

σy = Standard deviation of Y.

Karl Pearson's coefficient of correlation is also known as the product moment correlation coefficient. The value of 'r' lies between ±1. Positive values of r indicate positive correlation between the two variables (i.e.,

ISBN: 978-81-930585-3-4

changes in both variables take place in the same direction), whereas negative values of 'r' indicate negative correlation i.e., changes in the two variables taking place in the opposite directions. A zero value of 'r' indicates that there is no association between the two variables.

When r (+) 1, it indicates perfect positive correlation and when it is (-) 1, it indicates

perfect negative correlation, meaning thereby that variations in independent variable (X) explain 100% of the variations in the dependent variable (Y). We can also say that for a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction, then correlation will be termed as perfect positive. But if such change occurs in the opposite direction, the correlation will be termed as perfect negative. The value of 'r' nearer to + 1 or – 1 indicates high degree of correlation between the two variables.

- **Regression**

The correlation coefficient only expresses association and by itself tells us nothing about the causal relationships of the variates. Thus, purely from knowledge that two variates x and y are correlated, we cannot say whether variation in x is the cause or the results of the variation in y or whether the association results from mutual dependence of the two variates or from common causes affecting both of them.

Similarly, the more existence of a high value of correlation coefficient is not necessarily indicative of an underlying relationship between the two varieties.

The underlying relation between y and x in a bivariate population can be expressed in the form of a mathematical equation known as regression equation and is said to represent the regression of the variate y and the variate x (**Panse and Sukhatme**, 1967).

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.

$$\Sigma (y1 – 1 – bx1)2 \text{ with respect to a and b}$$

The values of a and b

$$a = y - bx$$

$$b = \Sigma xy - (\Sigma xy) (\Sigma yi) n \Sigma xi2 - (\Sigma xi)2n$$

The regression line can now be written as

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

Where b is the regression coefficient.

- **Multiple regression analysis**

When there are two or more than two independent variables, the analysis concerning relationship is known as multiple correlation and the equation describing such relationship as the multiple regression equation. We here explain multiple correlation and regression taking only two independent variables and one dependent variable (convenient computer programmes exist for dealing with a great number of variables). In this situation the results are interpreted as shown below :

Multiple regression equation assumes the form

$$\hat{Y} = a + b1X1 + b2X2$$

ISBN: 978-81-930585-3-4

Where X1 and X2 are two independent variables and Y being the dependent

variable, and the constants a, b1 and b2 can be solved by solving the following three

normal equations :

$$\Sigma = + \Sigma + \Sigma \; i \; i \; i \; Y \; na \; b \; X \; b \; X \; 1 \; 1 \; 2 \; 2$$

$$\Sigma = \Sigma + \Sigma + \Sigma \; i \; i \; i \; i \; i \; i \; X \; Y \; a \; X \; b \; X \; b \; X \; X \; 1 \; 2 \; 1 \; 2 \; 2 \; 1 \; 1 \; 1$$

$$\Sigma \; i \; i = \Sigma \; i + \Sigma \; i \; i + \Sigma \; i \; X \; Y \; a \; X \; b \; X \; X \; b \; X \; 2 \; 2 \; 2 \; 2 \; 1 \; 1 \; 2 \; 2$$

(It may be noted that the number of normal equations would depend upon the number of independent variables. If there are 2 independent variables, then 3 equations, if there are 3 independent variables then 4 equations and so on, are used).

In multiple regression analysis, the regression coefficients (viz., b1, b2) become less reliable as the degree of correlation between the independent variables (viz., X1, X2) increases. If there is a high degree of correlation between independent variables, we have a problem of what is commonly described as the problem of multicolinearity. In such a situation we should use only one set of the independent variable to make our estimate. In fact, adding a second variable, say X2, that is correlated with the first variable, say X1, distorts the value of the regression coefficients. Nevertheless, the prediction for the dependent variable can be made even when multicolinearity is present, but in such a situation enough care should be taken in selecting the independent variables to estimate a dependent variable so as to ensure that multicolinearity is reduced to the minimum.

With more than one independent variable, we may make a difference between the collective effect of the two independent variables and the individual effect of each of them taken separately. The collective effect is given by the coefficient of multiple correlations.

In multiple regressions we form a linear composite of explanatory variables in such a way that it has maximum correlation with a criterion variable. The main objective in using this technique is to predict the variability of the dependent variable based on its covariance with all the independent variables. One can predict the level of the dependent phenomenon through multiple regression analysis model, given the levels of independent variables.

- **Stepwise multiple regression**

Stepwise regression is a variation of multiple regression which provides a means of choosing independent variables that yield the best prediction possible with the fewest independent variables. It permits the user to solve a sequence of one or more multiple linear regression problems by stepwise application of the least square method. At each step in the analysis, a variable is added or removed which results in the greatest production in the error sum of squares (**Burroughs Corporation**, 1975).

According to **Draper** and **Smith** (1981), the method of stepwise multiple regression analysis is to insert variables in turn until the regression equation is satisfactory. The order of insertion is determined by suing the partial correlation coefficient as a measure of the importance of variables not yet in the equation. The programme, according to **Burroughs Corporation** (1975), first forms a correlation matrix, finds the best predictor (the independent variable having the highest correlation with the criterion variable) and performs a regression analysis with this predictor. Then, the second best predictor (independent variable having the second highest correlation with the criterion) is found and a regression analysis using the multiple correlations of the two best predictors is performed, and so on. At any given step, the group of predictors being used is not necessarily the best group of that size (i.e. the particular group of independent variables does not necessarily have the highest multiple correlation with the criterion that any group of this size does).

Rather, this group contains the variables that have the highest individual correlation with the criterion. Significance of a variable that is being considered for entrance into the regression equation is measured by the

ISBN: 978-81-930585-3-4

F-statistic. If F is too small (less that F 'include'), the variable is not added to the regression equation. *Include* statement establishes the minimum value of the F-statistic required for the inclusion of a variable in the regression equation. In the example which follows, the F-value for inclusion was 0.01.

Significance of variables already in the regression equation may change as new variables are entered. This significance of the variables currently in the equation is also measured by the F-statistic. If F is too small (less than F 'delete'), the variable is not added to the equation. Delete establishes the value of the F statistic below which the variable is deleted from the regression equation. Here the F-value for deletion was 0.005.

The 'tolerance' level specified is used as control of degeneracy. Degeneracy occurs when a variable entered into the equation is a linear combination of variables already present in the equation. Tolerance statement establishes the maximum value a pivoted element may attain while still allowing its associated variable to be brought into equation. A variable is not brought into the regression equation if its associated pivoted element is below the specified tolerance level, which was 0.001 in the present example.

- **Factor analysis**

Factor analysis is by far the most often used multivariate technique of research studies, especially pertaining to social and behavioural sciences. It is a technique applicable when there is a systematic interdependence in finding out something more fundamental or latent which creates this commonality. For instance, we might have data, say, about an individual's income, education, occupation and dwelling area and want to infer from these some factor (such as social class), which summarizes the commonality of all the said four variables. The techniques used for such purpose is generally described as factor analysis. Factor analysis, thus, seeks to resolve a large set of measured variables in terms of relatively few categories, known as factors. This techniques allows the researcher to group variables into factor (based on correlation between variables) and the factor so derived may be treated as new variables (often termed as latent variables) and their value derived by summing the values of the original variables which have been grouped into the factor. The meaning and name of such new variable is subjectively determined by the researcher. Since the factors happen to be linear combinations of data, the coordinates of each observation or variable is measured to obtain what are called factor loadings. Such factor loadings represent the correlation between the particular variables and the factor, and are usually placed in a matrix of correlations between the variable and the factors.

- **Principal component analysis**

There are several methods of factor analysis. The method of Principal Component Analysis which is widely used is discussed here. The principal component analysis extracts m-eigenvectors (principal component axes) and corresponding m-eigenvalues (the variance measured along the eigenvector), from m x m symmetrical matrix of correlation. The eigenvectors obtained from this principal component analysis are all orthogonal (i.e. inter-column correlations are near zero).

The eigenvalues account for all of the original data variances in decreasing order such that each has variance or eigenvalue less than the previous ones. The total of the eigenvalues $\lambda 1 + \lambda 2 + \ldots\ldots\ldots + \lambda m$ which is the same as the sum of the variances constituting the diagonal or trace of the correlation matrix before transformation. The principal components are then converted into factors by multiplying each element of the principal components or eigenvectors (V) by the square-root of the corresponding eigenvalues (λ ½. V). Factors, thus, besides the direction also represent the variances.

The analysis calls for the selection of a minimum number of meaningful and useful factors, considerably fewer in number than the original variables, which will account for most of the variances in the data set and, therefore, convey the same information. Various criteria for selection of suitable factors are available. **Kaiser (**1958) and others have recommended retaining all those eigenvalues which have values more than 1.

Next step is to remove the noise imposed by (m - p) unnecessary axes. To accomplish this, p-orthogonal reference axes or factors are orated about the origin to positions such that the variance of the loading from

ISBN: 978-81-930585-3-4

each variable onto each factor axis is either extreme (± 1) or near zero. This maximistion of the range of the loadings was performed by using Kaiser's Varimax criterion. Scanning through each factor column for large absolute values in the varimax matrix will reveal a few variables with significantly high loadings and many others with insignificant loadings. The column showing communality (Σ h2j) is the total amount of variance of each variable retained in the elements of the factors in each row of the varimax matrix. Fairly high communality of each variable implies the appropriateness of the model adopted, for the study. The last step involved meaningful interpretation of the factors.

- **Path analysis**

The term "Path Analysis" was first introduced by the 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 casual 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 inter relationship between father's education, father's occupation, son's education, 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 to a series of 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 casual connections and then by elucidating the indirect effect of the explanatory variables.

- **Cannonical Correlation Analysis**

Canonical correlation analysis is used to identify and measure the associations among two sets of variables. Canonical correlation is appropriate in the same situations where multiple regression would be, but where are there are multiple intercorrelated outcome variables. Canonical correlation analysis determines a set of canonical variates, orthogonal linear combinations of the variables within each set that best explain the variability both within and between sets.

ISBN: 978-81-930585-3-4