Chapter 5

Research Methodology

In this chapter 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. The entire discussion for easy understanding has been made under the following sub-heads.

A. Locale of research
B. Sampling design
C. Pilot Study
D. Variables and their measurements
E. Methods of data collection
F. Statistical tools used for analysis of data analysis.
G. Participatory exercises
H. SWOT analysis

A. Locale of Research:

The present study was conducted in Haringhata block of Nadia district, Ausgram I block of Burdwan district and Habra II block of North 24 Parganas district of the state West Bengal were selected for the study.

- The characters and the factors under study have been well discernible to this area;

ISBN: 978-93-85822-62-9

- The researcher's close familiarity with respect to area, people, officials and local dialects;
- The ample opportunity to generate relevant data due to the close proximity of the area with the research and extension wing of the state Agriculture;
- The highly cooperative, responsive respondents;
- The prevalence of the characters under study has been observed to get relevant information;
- Experienced, well versed, venturesome and risk bearing farm entrepreneurs;
- Easy accessibility of the area;

B. Sampling design:

Purposive sampling techniques were adopted for the study. For selection of state, district, block, purposive sampling techniques was adopted because the area was ideal for entrepreneurship study, convenient for researcher to access and having the infrastructural facilities and also respondents are purposively selected.

C. Pilot Study:

Before taking up actual fieldwork a pilot study was conducted to understand the area, its people, institution, communication and extension system and the knowledge, perception and attitude of the people towards farm entrepreneurship concept.

Sampling Technique and Sampling Design

Step	Items	Level			Approach
1.	State	West Bengal			Purposive
2.	Districts	Burdwan	Nadia	North 24 Parganas	Purposive

3.	Subdivision	Bardhaman Sadar North subdivision	Kalyani subdivision	Barasat Sadar subdivision	Purposive
4.	Block	Ausgram-I	Haringhata	Habra -II	Purposive
7.	Respondents	16	15	24	Purposive
Total No. of Respondents: 55					

D. Variables and their measurements

After reviewing various literature related to the field of study and consultation with the respected chairman of Advisory Committee and other experts, a list of variables was prepared. On the basis of the selected variables, a schedule was constructed.

I) Independent Variables:

Sl. No.	Variables	Notation	Score
1	Age	X_1	Chronological age
2	Participation Index	X_2	Total no of meeting attained
3	Education	X_3	Year of Schooling
4	Marital Status	X_4	Period of marriage
5	Family Size	X_5	Number of family members
6	Land Holding	X_6	Bigha (Homestead + cultivable)
7	Cropping Intensity	X_7	(Gross cropped area/Net cropped area)*100
8	Income	X_8	Family income/Year in Rs.
9	Expenditure	X_9	Family expenditure/Year in Rs
10	Distance	X_{10}	In km
11	Conveyance	X_{11}	In Rs

i. **Age (X_1):**In all societies, age is one of the most important determinants of social status and social role of the individual. It is said that young farmers are more inclined to the change than the aged farmer who usually stick to their traditionally bound old practices. In the present study, age of the respondent was measured on the basis of their chronological age at the time of investigation.

ii. **Participation Index(X_2):** Participation is the the action of taking part in something. Participation index is scored astotal no of meeting attained.

iii. **Education (X_3)**: Education is instrumental in building personality structure and helps in changing one's behaviour in social life. Education may be conceptualized as the amount of formal schooling attained/ literacy acquired by the responded.

iv. **Marital Status(X_4):**Marital status, is any of several distinct options that describe a person's relationship with a significant other. marital status are terms used in forms, vital records, and other documents to ask or indicate whether a person is married or single. In the simplest contexts, no further distinction is made. A status of married means that a person was wed in a manner legally recognized by their jurisdiction. A person's specified civil status might also be married if they are in a civil union or marriage. It is scored as period of marriage in year.

v. **Family size (X_5)**: Total number of adult and minor family member present in a family.

vi. **Land Holding (X_6):** The amount of land owned by a person is an important parameter to assess the economic status of the person in the society. The attribute size of holding had been measured by addition of homestead and cultivable land in Bigha.

vii. **Cropping Intensity (X_7):** It is calculated by (Gross cropped area/Net cropped area)*100

viii. **Income (X_8):**The Annual Income of a person is an important parameter to assess the economic status of the person in the society. Total income from farm and off-farm sources earned by the unit of family in a year. Annual income has been scored as family income/year in Rs.

ix. **Expenditure (X_9):** The Annual expenditure of a person is an important parameter to assess the economic status of the person in

ISBN: 978-93-85822-62-9

the society. Total expenditure from health,education, farming and other sources by the unit of family in a year. Annual expenditure has been scored as Family expenditure/Year in Rs.

x. **Distance (X_{10}):** Distance from ATMA office has been recorded in km.

xi. **Conveyance(X_{11}):** Amountof money spend in conveyance is recorded in Rs.

II) Dependent Variables:

1.	Perception Level (Y_1)	Perception level of respondent about ATMA is calculated by the questions depicted in the schedule and measured in %
2.	Impact of ATMA (Y_2)	Measured in %
3.	No of Projects (Y_3)	Total no of project operated in a year
4.	Fund Utilization Efficiency (Y_4)	(Fund utilized /fund sanctioned)*100
5.	Success of The Project (Y_5)	Measured in %

E. Methods of Data Collection:

a) Preparation of Interview Schedule:

On the basis of the findings of pilot study a preliminary interview schedule was formed with the help of literature review and by the assistance of Chairman of Advisory Committee. The interview schedule consisted of six major parts according to the objectives of the study.

b) Pre-testing of Interview Schedule:

Pretesting or preliminary testing is the process of an advance testing of the study design after the schedule/questionnaire has been prepared. The object of pretesting is to detect the discrepancies that have emerged and to remove them after necessary modification in the schedule. It also helps to identify whether the questions are logically organized, the replies could properly recorded in the space provided for or there is any scope for further improvement. After conducting pretesting appropriate changes and modification of the interview schedule have been made. The individuals who

responded in pretesting have been excluded in the final sample selected for the study.

c) Techniques of field data collection:

The respondents were personally interviewed. The items were asked in Bengali as well as English version in a simple terminology so that the members could understand easily. The entries were done in the schedule by student investigator as well as respondent at the time of interview.

F. Statistical tools used for analysis of data :

a) Mean
b) Standard deviation
c) Coefficient of Variance
d) Correlation of coefficient
e) Multiple regression analysis
f) Path analysis
g) Canonical covariate analysis

a) 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 simplest and relatively stable measure of central tendency. The mean reflects and is affected by every score in the distribution.

When the data are expressed in a frequency distribution (grouped), the mean is calculated by using the following formula–

$$X = \frac{\sum_{i=1}^{N} fixi}{N}$$

Where,

x = Mean of the observation.

fi = Frequency of the class.

xi = Mid value of the class.

N = Total number of observation

b) Standard deviation:

Standard deviation (SD) of a set of observation is the square root of the arithmetic mean of the squares of the deviations. The deviations being measured from the arithmetic mean of the distributions. It is commonly denoted by the symbol (Sigma). To measure the average deviation from the standard value of the data standard deviation is used. It is less affected by sampling errors and is a more stable measure of dispersion.

The standard deviation of the data grouped in the form of frequency distribution is computed by using the following formula–

$$S.D. = \sqrt{\frac{\sum_{i=1}^{N} fixi^2}{N} - \left[\frac{\sum_{i=1}^{N} fixi}{N}\right]^2}$$

Where,

= Standard deviation

N = total No of observation in a particular cell.

X = value of observation in a particular cell

F = Frequency of observation

$\overline{X}$= mean number of observation

i = any number (e.g. 1, 2, 3) denoting position

c) Coefficient of Variance:

A measure of variation which is independent of the unit of measurement is proved 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 percentage of the mean.

Coefficient of variation is measured by using the following formula –

$$C.V. = \frac{S.D.}{Mean} \times 100$$

d) Coefficient of correlation:

When an increase or decrease in one variety is accompanied by an increase or decrease in another 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 rational level of measurement and are linearly related. A Pearson 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 scares 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
∑X²	=	Each of X squared, then summed
(∑X) ²	=	Sum of X score squared
∑Y	=	Sum of Y scores

$\sum Y^2$ = Each of Y squared, than summed

$(\sum Y)^2$ = Sum of Y score squared

The range of correlation coefficient is between -1 to +1. This means that -1 is perfect negative correlation and +1 is perfect positive correlation. A perfect correlation is, however, seldom achieved. A correlation coefficient to be acceptable should be statistically significant. Otherwise, we say that no significant relationship exist between the variables.

e) Multiple regression analysis:

Generally a number of antecedent variables simultaneously contribute to influence the consequent variables, as in the case under study. It is of immense practical value to know the extent to which the antecedent variables, individually or jointly, could predict or contribute towards the consequent variable. This was done by computing multiple regression analysis. If Y is the consequent variable and X1, X2, X3are the antecedent variables; the multiple regression equation is given by the following formula-

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

Or, $Y = a + \sum bx$

The significance of the b- value was judged by calculating their respective t- values and comparing them to the table, given by Fisher and Yates (1963), with (n-p-1) degree of freedom (where, n = number respondents and p = number of antecedent variables) at 5% and 1% level of significance.

The square root of the ratio of the regression sum of squares to the total sum of squares is known as multiple correlation coefficients and is denoted by R. The square of the multiple correlation coefficients R2 is called the multiple coefficient of determination and represents the fraction

of the variation of Y accounted for by its joint association with the variables X1, X2, X3…

Central to the application of multiple regression analysis is the interpretation of the final fitted model. A significant F- value for R means that the fitted model is adequate. The significance of the F- value was judged by comparing it to the table value, given by Fisher and Yates (1963) with P and (n-p-1) degrees of freedom (where, P = number of antecedent variables and n = number of respondents) at 5% and 1% levels.

Stepwise multiple regression: Stepwise regression is a variation of multiple regressions 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).

f) Path Analysis:

The objective of doing Path Analysis is to get a clear picture of the direct and indirect effects of the independent variables on the dependent variable. Variables, through which substantial indirect effects are channelled, are also found out.

Singh and Chaudhary (1977), defined path coefficient as the ratio of the standard deviation of the effect due to a given cause to the total standard deviation of the effect i.e. if Y is the effect and x1 is the cause, the path coefficient for the path from cause x1 to the effect Y is $x1/y$

It is advisable to do path analysis with only those variables which have significant effects on the dependent variable. This may be done by restoring

to multiple regression analysis, and selecting those independent variables whose partial b value are significant. This shall enhance clarity of the path analysis.

g) Canonical correlation analysis:

In statistics, canonical-correlation analysis (CCA) is a way of making sense of cross-covariance matrices. If we have two vectors X = (X1, ..., Xn) and Y = (Y1, ..., Ym) of random variables, and there are correlations among the variables, then canonical-correlation analysis will find linear combinations of the Xi and Yj which have maximum correlation with each other. Virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical-correlation analysis, which is the general procedure for investigating the relationships between two sets of variables. The method was first introduced by Harold Hotelling in 1936.

Given two column vectors $X = (x_1, \ldots, x_n)'$ and $Y = (y_1, \ldots, y_m)'$ of random variables with finite second moments, one may define the cross-covariance $\Sigma_{XY} = \operatorname{cov}(X, Y)$ to be the $n \times m$ matrix whose (i,j) entry is the covariance $\operatorname{cov}(x_i, y_j)$. In practice, we would estimate the covariance matrix based on sampled data from X and Y (i.e. from a pair of data matrices).

Canonical-correlation analysis seeks vectors a' and b' such that the random variables $a'X$ and $b'Y$ maximize the correlation $\rho = \operatorname{corr}(a'X, b'Y)$. The random variables $U = a'X$ and $V = b'Y$ are the first pair of canonical variables. Then one seeks vectors maximizing the same correlation subject to the constraint that they are to be uncorrelated with the first pair of canonical variables; this gives the second pair of canonical variables. This procedure may be continued up to $\min\{m, n\}$ times.

G. Participatory exercises with the help of PRA tools:

1. Time trend analysis:

Time-trend designs are a form of longitudinal ecological study, and can provide a dynamic view of changes of data over years.

2. Problem Cause and solution analysis:

By this study we can identify the problems and their causes and give some solutions of the problem.

3. Matrix ranking: Selection of an issue to investigate with respondent.

- A serial number should be given to each product or service
- Respondent asked to give score in order of performance against each criteria
- Matrix scoring is a variation of matrix ranking. Instead of ranking each item against each criteria, score them out of 10.

4. Venn diagram:

It is used to depict key institutions, organizations and individuals and their relationship with the local community or others. Key players in decision making are shown. On the Venn diagram each institution is represented by a circle. The size of the circle represents the importance, significance or power of that institutions.

H.SWOT analysis:

SWOTanalysis (alternatively SWOTmatrix)isan acronym for strengths, weaknes ses, opportunities, and threats and is a structured planning method that evaluates those four elements of an organization, project or business venture. A SWOT analysis can be carried out for a company, product, place, industry, or person. It involves specifying the objective of the business venture or project and identifying the internal and external

factors that are favorable and unfavorable to achieve that objective. Some authors credit SWOT to Albert Humphrey, who led a convention at the Stanford Research Institute (now SRI International) in the 1960s and 1970s using data from Fortune 500 companies.[1][2] However, Humphrey himself did not claim the creation of SWOT, and the origins remain obscure. The degree to which the internal environment of the firm matches with the external environment is expressed by the concept of strategic fit.

- Strengths: characteristics of the business or project that give it an advantage over others
- Weaknesses: characteristics of the business that place the business or project at a disadvantage relative to others
- Opportunities: elements in the environment that the business or project could exploit to its advantage
- Threats: elements in the environment that could cause trouble for the business or project

ISBN: 978-93-85822-62-9