Chapter–4

The Empirical Approach and 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.

4.1. Locale of research
4.2. Sampling design
4.3. Pilot Study
4.4. Variables and their measurements
4.5. Methods of data collection
4.6. Statistical tools used for analysis of data.

4.1. Locale Of Research

The present study has been conducted in two adjoining states, West Bengal and Bihar. The village, Ghoragachha of chakdah block in Nadia district of the state West Bengal, and the village, Chiroura of Naubatpur block in Patna district of the state Bihar have been selected for the study.

- The characters and the factors under study have been well discernible to this area;
- 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 profuse scope to get relevant information regarding discontinuance, disagreement, conflict, rejection, dissonance, reinvention and confusion regarding agricultural technology;
- Experienced, well versed, venturesome and risk bearing farm entrepreneurs;
- Easy accessibility of the area;
- The study would help the researcher to conduct diversified extension programs and activities in future.

4.2. Sampling Design

The purposive as well as simple random sampling techniques have been adopted for the present study. It may be termed as multistage random sampling procedure. The two states namely West Bengal and Bihar has been considered purposively. The two districts one in West Bengal, named Nadia and another in Bihar, named Patna has been selected purposively for the study. Two blocks one in district Nadia, named Chakdah and another one Naubatpur from Patna district has been selected purposively for the study. Two villages one in block Chakdah, named Ghoragachha and another in Naubatpur block, named Chiroura has been selected purposively for the study. 150 respondents, 75 from each village had been selected randomly for finale data collection.

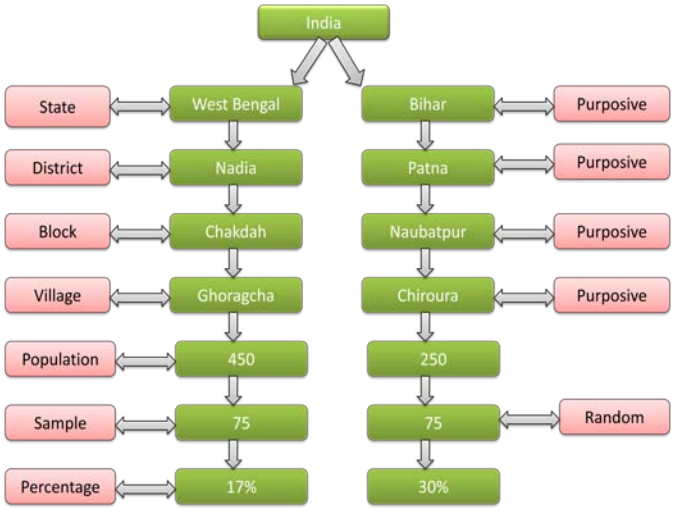

Fig. 4.1: Sampling Design

4.3. Pilot Study

A pilot study has been conducted in the selected villages before constructing the data collecting devices. In course of this survey, informal discussion has carried out with some farmers, local leaders and extension agents of the localities. An outline of the socio-economic background of the farmers of the concerned villages, their opinion towards different types of technology socialization process, innovation-decision process, discontinuance, disagreement, conflict, rejection, dissonance, reinvention and confusion helped in the construction of reformative working tools.

The components of pilot study are:

- General information;
- Specific information;
- Prevalence of variables;
- Body languages of the prospective respondents;
- Access to physical location;

ISBN: 978-81-930585-0-3

- The type, level and intensity of responsiveness;
- Related information including Agriculture.

4.4 Variables and Their Measurements

Variables in the present study have been categorized into two main categories.

- Independent variables;
- Dependent variables.

The Independent Variables

Age (x_1)

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 of the respondents lived since birth at the time of interview, has been taken as a measure of age of the farmer.

Education (x_2)

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

Family Education Status (x_3)

In the present study family education status is the highest education achieved by the members of the family.

Educational Aspiration (x_4)

The educational aspiration is operationalized as the respondent desire to educate his/her sons and daughter. The respondents were therefore, asked to simply mention the level up to which they desired to educate their children.

Family size (x_5)

Family size is opretionalized as the members in the individual family. In the present study only those members of the family considered, who were taking the meal in one chullah.

Gender (x_6)

Gender is operationalized in the present study as number of male in family divided by the number of female in the same family.

Urbanization Index (x_7)

Urbanization Index indicates transformation from rural economies to modern industrial one. Urbanization is the product of demographic explosion and poverty induced rural-urban migration. In the present study schedules have been developed as follows.

Watching T.V. in hrs
Using Toilet soap: score was mentioned as "1" for yes and "0" for No.
Consulting Doctors (Frequencies) in Years
Possessing Bike/Cycle: score was mentioned "1" for cycle and "2" for Bike.
Visiting city centers: frequency in year
Scores for all the item has been summed up and then scores obtained has been divided by total number of items.

Occupation (x_8)

Occupation of a person refers to regular activity performed for payment that occupies one's time. In the present study a scale is developed on 1-6 point scale. This scale consists of as follows.

Sl. No.	Items	Scale (1-6)
1	Labor	
2	Artisan	
3	Business	
4	Independent	
5	Farming	
6	Services	

Cropping Intensity (x_9)

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

$$\frac{Total\ annual\ cropped\ area}{size\ of\ holding} \times 100\ \%$$

Farm size (x_{10}) Farm size is a measure of farm business. Operationally farm size may be defined as a tract of land possessed by an individual for the purpose of growing crops. Different research workers had tried to measure farms size in different ways. In the present study, actual area under cultivation in bigha is divided by size of the family taken as measure of farm size.

$$\frac{Actual\ farm\ holding}{size\ of\ the\ family}$$

Expenditure Allotment (x_{11})

In the present study per cent of expenditure incurred on farming of the total expenditure incurred annually by farm family. Total expenditure has been calculated as follows

- Expenditure incurred on food annually.
- Expenditure incurred on Clothes annually.
- Expenditure incurred on Education annually.
- Expenditure incurred on Farming annually.
- Expenditure incurred on health annually.

ISBN: 978-81-930585-0-3

$$\frac{Expenditure\ \ incurred\ \ on\ farming\ \ annually}{Total\ \ Expenditure} \times 100$$

Credit load (x_{12})

Credit load of the farmers indicates that how much credit farmers have in outstanding. In the present study credit load of the farmers has been calculated as

$$\frac{Credit\ \ outstanding}{Size\ \ of\ \ the\ \ family}$$

Annual Income (x_{13})

Annual income is the economic measurement of farmers' status. It is operationally defined as the gross income from all the viable sources of income in a single year. It is measured in terms of rounded of rupees. The gross income is constituted by the total income generated from agriculture, dairy, poultry, fishery enterprises, business and services. In the present study it has been calculated with the formula as follows.

$$\frac{Total\ \ Income\ \ in\ a\ year}{Family\ \ size}$$

Electricity consumption (x_{14})

Electricity consumption is calculated in the present study with the formula as follows.

$$\frac{Electricity\ \ consumption\ \ in\ a\ mont\ h\ of\ \ a\ farm\ \ family}{Size\ \ of\ \ the\ family}$$

Fuel consumption (x_{15})

Fuel consumption shows that diesel, petrol and kerosene oil consumed by a farm family in a year. In the present study it has been calculated as follows

$$\frac{Consumption\ \ \ of\ diesel\ ,petrol\ \ and\ kerosene\ \ oil\ in\ a\ year}{Size\ \ of\ \ the\ family}$$

Irrigation Index (x_{16})

In the present study irrigation index is calculated in per cent as follows

$$\frac{Area\ \ of\ \ the\ land\ \ under\ \ Irrigaion}{Total\ \ holding\ \ size} \times 100$$

Adoption leadership (x_{17})

Adoption leadership operationalized as the degree to which an individual initiates or motivates the action of the other fellows to adopt new ideas.

In the present study, adoption behavior of the framers has been measured with the help of the scale, developed by Maulik (1965) with modification. He defined adoption leadership as the personality characteristic of a farmer revealing the degree to which he is an adoption leader. The eight statements constituted the total scale with the weightage of 10 (1-10). The summation of the total response scores divided by total no. statements (All together 8 statements) revealed the total score of adoption leadership of the farmers.

Scientific orientation (x_{18})

In the present study scientific orientation is operaionalized as the characteristic of individual, which made him to trust and rely on ideas and practices developed through scientific research. This variable is measured with the help of scale, developed by Supe (1969) with modification. The scale consisted of five statements and each statement is of 10 (1-10) point scale and asked to the farmer to give preference score out of 10 (1-10). The score for each respondent in scientific orientation scale has been obtained by summating the score for each statement divided by total statement (all together 5 statements).

Independency (x_{19})

In the present study independency has been operationalized as the extent of feeling of one's own ability and resourcefulness in carrying out any activity. It has been measured with the help of independence scale developed by Supe (1969) with some modification. The scale consisted of 5 statements and each statement is of 10 (1-10) point scale and asked to the farmers to give preference score out of 10 (1-10). The score of each individual in the independency has been obtained by summing the score for each statement divided by total statement (all together 5 statements).

Innovation proneness (x_{20})

Innovation proneness indicates the behavior pattern of individual who have interest in and desire to seek change in farming techniques and to introduce such change into their operations wherever practical and feasible.

In the present study, Maulik's (1965) with modification self-rating innovation proneness scale has been used to measure the innovation proneness of the farmers. This scale consisted of three set of statements. Each set of statement contained three short statements with weightage of 10 (1-10) point each in every set of statement of innovation proneness and farmers preference has been asked to give score out of 10 (1-10) for every statement in each section. The forced choice method has been followed to overcome the familiar problems of personal bias and lack of objectivity in self-evaluation. This method forced the respondents to choose from the group of three short statements describing a particular personality characteristic; the one most accurately portrayed the respondent himself. After obtaining the respondent's score out of 10 (1-10) choices for each of the three sets of the statements, the scoring has been done by summing up score of all statement divided by

total no. of statement (all together 9 statements). As there were three sets of statements, for innovation proneness scale, the sum of score of all statements divided by total no. of statement is a respondents' self-rating score for innovation proneness.

Risk orientation (x_{21})

In the present study, the risk orientation of farmers has been measured with the help of risk preference scale developed by Supe (1969) with modification. Supe, defined risk preference as the degree to which a farmer was oriented towards risk and uncertainty and had the courage to face the problems in farming. The scale consisted of six items. The items are rated in out of 10 point scale and farmers preferences have been asked to give score out of 10.

Economic motivation (x_{22})

Economic motivation has been conceptualized as the values or attitudes, which attach greater importance to profit maximization as the ends and means. In this study, economic motivation of the farmers has been measured with the help of the self-rating economic motivation scale developed by Maulik (1965) with some modification. The scale consists of three sets of short statements, having total 8 statements with weightage of 10 point scale in every set of statement for economic motivation. The respondents out of 10 point scale choices for each of the three sets of statements have been recorded.

Orientation towards competition(x_{23})

Singh (1981) defined the variable as the degree to which a farmer is oriented to place himself in a competitive situation in relation to other farmers for projecting his excellence in farming.

Sing (1981) with modification scale has been used to get the score for the orientation towards competition in out of 10 (1-10) point for each of the statements. The score for each individual in the orientation towards competition have been obtained by summing up the scores of all statement divided by total no. of statement (All together 6 statements).

Planning orientation (x_{24})

Samanta (1977) defined management orientation as the degree to which a farmer is oriented towards scientific farm management comprising of planning, production and marketing functions in his farm.

Management orientation is operationalized as the degree to which the individuals are oriented towards managing their income generating activities regarding planning, production and marketing function.

The management orientation scale developed by Samanta (1977) with modification was adopted for this study, consisted of 6 statements. The statements have benn rated in out of 10 (1-10) point scale. The scoring has been done by summing up score of all statements divided by total no. of statements (All together 6 statements).

Production orientation (x_{25})

Scale used as same as Planning orientation.

Market orientation (x_{26})

Scale used as same as Management orientation.

Social participation(x_{27})

In the present study social participation has been operationalized as the respondent's interaction with the organization. A suitable scale has been developed, which consisted of total 6 statements regarding their membership in the social organization and interaction with the society. The scale consists of 6 statements with weightage of 10 (1-10) points. The respondents out of 10 (1-10) point scale choices for each of the three sets of statements have been recorded. The scoring has been done by summing up scores of all statements, and, the resultant sum has been divided by total number of statement (All together 6 statements).

Utilization of cosmopolite source of information (x_{28})

The utilization of sources of information refers to the use of information sources both from outside and within the community for getting agricultural information by the farmer.

The sources of information of a farmer have been determined by summing up the sources obtained from the cosmopolite and localite schedule and dividing by total no of items (All together 23 statements item) developed by Singh (1993) with some modification adopted for this study.

The cosmopolite sources schedule consisted of 23 statements. The statements have been rated in four-point response categories: 'Very often', 'Often', 'Some times', and 'Never'. The statements have been given scores for

Statements	:	Score
Very often	:	4
Often	:	3
Sometimes	:	2
Never	:	1

The localite sources of information, the farmer have been asked to indicate, with, whom the following personal, he has used as his sources of information in the previous: 'local agent', 'local leaders', 'friends/neighbours/relatives' and

ISBN: 978-81-930585-0-3

'progressive/experienced farmers'. Each respondent has been asked to rate them on four point response categories: 'very often'–'4', 'often'- '3', 'Sometimes'- '2' and 'Never'- '1'

Information Seeking Behavior (x_{29})

Information seeking is a kind of behavior of respondent. International Encyclopedia of Information and Library Science defines Information seeking behavior as "The complicated form of actions, which people slot in, when seeking information of whatever kind for whatever reason". (David Ellis, 2003).

In the present study a suitable scale has been developed which consisted of total 7 statements. The statements has been rated in out of 10 (1-10) point scale response categories and farmers' preference has been asked to give score out of 10 (1-10). The score of the individual has been obtained by summing up the scores of the 7 statements divided by total no. of items (All together 7 statements).

Training received (x_{30})

In this study, training received refers to whether an individual farmer has attended the farmers training programs or not during last three years (from the date of interview) and so, for how many days. From the farmers training programs, farmers generally gather some new knowledge regarding production and infrastructure available. This measurement was previously done by De and Rao, (2000).

It may be assumed that farmers training programs is one of the means by which the desired changes in knowledge and skills of farmers of farmers can be brought out. The individuals' attitude and motivation, influenced by training programs will lead to their adoption behavior.

In the present study, the attendance of respondents in the training meetings in number of days multiplied by their educational score plus one has been taken as a measure of their training received score.

Distance matrix (x_{31})

In the present study distance matrix is the mean distance covered by the respondents for selling their crop, for getting credit from the bank, for getting treatment of their ailments or diseases and for sending their children to school and colleges for education.

Drudgeries (x_{32})

Use of human power is extensive in several farm operations such as handling of farm machinery and tools, various agricultural activities *viz*. ploughing, sowing, irrigating, spraying of pesticides, fertilizer application, monitoring. In the present study Drudgeries is the work load faced by the farmers in hours. It has been operationalized as the total number of hours spent in major agricultural activities or livelihood activities divided by number of items.

Dependent Variables

The appropriate operationlization 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 has given insight into the contemplation part of the farmer's psyche, who is dealing with the agricultural innovation. It has been considered here that the entire post-adoption phenomenons are under a single continuous contemplating process. For this reason the measurement of these variables has been carried out in following manner.

- Perception on Discontinuance (y_1)
- Perception on Rejection (y_2)
- Disagreement (y_3)
- Conflict (y_4)
- Reasons for Dissonance (y_5)
- Reasons for Reinvention (y_6)
- Confusion Index (y_7)
- Social Entropy (Y)

The dependent variables *viz.* perception on Discontinuance (y_1), Perception on Rejection (y_2), Disagreement (y_3), Conflict (y_4), Reasons for Dissonance (y_5), Reasons for Reinvention (y_6), Confusion Index (y_7) has been measured using structured interview schedule and questionnaire method and responses have been measured on (1-10) point scale. An appropriate questionnaire has been developed in the following manner.

Dependent Variables	Empirical Measurement
Perception on discontinuance (y1)	Items developed, judges rating done against 10 point scale and significant statements were screened out.
Perception on rejection (y2)	Items developed, judges rating done against 10 point scale and significant statements were screened out.
Disagreement (y3)	Items developed, judges rating done against 10 point scale and significant statements were screened out.
Conflict (y4)	Items developed, judges rating done against 10 point scale and significant statements were screened out.
Reasons for dissonance (y5)	Items developed, judges rating done against 10 point scale and significant statements were screened out.
Reasons for reinvention (y6)	Items developed, judges rating done against 10 point scale and significant statements were screened out.

ISBN: 978-81-930585-0-3

Confusion index (y7)	Items developed, judges rating done against 10 point scale and significant statements were screened out.
Social entropy (Y)	Multiplied value of the measuring items.

Social entropy: Social entropy has been conceived as the resultant disposition of interaction amongst the constituents' psycho-motivational characters. In the present study Social entropy is the multiplication of Perception on discontinuance (y_1), Perception on rejection (y_2), Disagreement (y_3), Conflict (y_4), Reasons for dissonance (y_5), Reasons for reinvention (y_6) and Confusion index (y_7)

$$\text{Social Entropy (Y)} = \frac{y1 \times y2 \times y3 \times y4 \times y5 \times y6 \times y7}{7}$$

4.5. Methods of Data Collection

4.5.1 Construction Of Schedule After Pre-Testing

The draft schedule for collection of data, incorporating the tools and techniques of different variables has been presented twice, each time on contact farmers. The quantification has been done for each and every variable after operationalizing them. Before final data collection, entire schedule has been pretested for elimination, addition and alteration with non-sample respondents of the study area. In pre-testing, care has been taken not to include respondents who are selected as sample for final interview. On the basis of experience in pre-testing, appropriate changes in the construction of item and their sequence have been made. The schedule has been then finalized and multiplicated. The final form of the schedule is given in the appendix.

4.5.2 Field Data Collection

The primary data in the present study has been collected directly from the farmers with the help of structured schedule through personal interview methods. Only the functional head of the household has been taken as respondents for the study. The personal interview method has been followed during the month of November 2011 to April 2012 to collect the relevant information form targeted respondents. In each village, before starting the interview, a few days have been devoted to establish rapport with the respondents. The schedule has been then administered to the respondents in local language. The responses from the farmers slected have been recorded in English on the schedule. The interview has been carried out by the researcher himself.

4.5.3 Method Of Secondary Data Collection

The case study method has been followed from the qualitative analysis and involves a careful and complete observation of a social unit, be that unit a person, a family, an institution, a cultural group or even the entire community. It is a method of study in depth rather than breadth. The case study places more emphasis on the full study deals with the process that take place and their interrelationship. Thus case study is essentially an intensive investigation of the particular unit under consideration. The object of the case study method is to locate the factors, which account for the behavior patterns of the given unit as an integrated totality.

1. CHARACTERISTICS OF CASE STUDY

The important characteristics of the case study method are as as follows;

a) The researcher can take one single social units for his study purpose; he may even take a situation to study the same comprehensively.

b) Here the selected unit is studied intensively i.e., it is studied in minute details. Generally, the study extends over a long period of time to ascertain the natural history of the unit so as to obtain enough information for drawing correct inferences.

c) In the context of this method we make complete study of the social unit covering all facets. Through this method we try to understand the complex of factors that are operative within a social unit as an integrated totality.

d) This approach is qualitative rather than quantitative. Every possible effort is made to collect information concerning all aspects of life. As such, case study deepens our perception and gives us a clear insight into life. For instances, this method considers not only how many crimes a man has done but shall also consider the factors that forced him to commit crimes when we are making a case study of a man as a criminal. The objective of the study may be to suggest ways to reform the criminal.

e) In respect of the case study method an effort is made to known the mutual inter-relationship of causal factor.

f) In this method the behavior pattern of the concerning unit is studied directly and not by an indirect and abstract approach.

g) Case study method results in fruitful hypothesis along with the data, which may be helpful in testing them, and thus it enables the generalized knowledge to get richer and richer. In its absence, generalized social science may get handicapped.

Assumptions

The case study method is based on several assumptions; the important assumptions may be listed as follows

a) The assumption of uniformity in the basic human nature in spite of the fact that human behavior may vary according to situations.

b) The assumptions of studying the natural history of the unit concerned.
c) The assumption of comprehensive study of the unit concerned.

Keeping all the assumptions in mind the researcher had followed or carried out the case study method to collect the secondary data in a view to get a complete study in generalized form of the attitudinal phenomena like adoption, rejection and discontinuance.

4.6 Statistical Analysis and Interpretation of Data (Analytical Tools)

The data collected in the interview schedule have been 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 Statistical techniques and tools used in the present study.

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 follows

$$\text{Mean or } (\bar{x}) = \frac{\sum x_i}{N} = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots\ldots\ldots\ldots x_n}{N}$$

Where,

($\bar{x}$)=The symbol we use for mean (pronounced as x bar)
$\sum$=Symbol for summation
$x_{i=}$Value of the i^{th} 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 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 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 average, especially in economic and social studies where direct quantitative measurements are possible.

Standard Deviation

Standard deviation is the most widely used measure of dispersion of a series and is commonly denoted by the symbol **σ** (pronounced as sigma) Standard deviation is the square root of the arithmetic mean of the square of the deviations, the deviations being measured from the arithmetic mean of distribution. It is less affected by sampling errors and is more stable measure of dispersion. It is worked out as follows,

$$\text{Standard deviation } (\boldsymbol{\sigma}) = \sqrt{\frac{\sum (x_i - x)^2}{n}}$$

Coefficient of Variation

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

$$\text{CV} = \frac{Standard\ \ Deviation\ \ (\sigma)}{Mean} \times 100$$

Coefficient of Correlation

When an increase or decrease in one variable is accompanied by an increase or decrease in other variable, 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 or measurement and are linearly related. A Karl Pearson's coefficient of correlation also known as 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
$\sum xy$=Each x multiplied by its corresponding y, then summed
$\sum x$ =Sum of x scores
$\sum x^2$ =Each x squared, then summed
$(\sum x)^2$ =Sum of x scores, squared
$\sum y$ =Sum of y scores
$\sum y^2$ =each y squared, then summed
$(\sum y)^2$ =Sum of y scores, squared

This coefficient assumes the following;

- That there is linear relationship between the two variables;
- That the two variables are causally related which means that one of the variable is independent and other is dependent and;
- A large number of independent causes are operating in both variables so as to produce a normal distribution.

The value of 'r' lies between +1 to -1. Positive values of r indicate that positive correlation between the two variables (i.e. changes in both variables take place in the same

ISBN: 978-81-930585-0-3

direction), whereas negative values of 'r' indicate negative correlation i.e. changes in the two variables taking place in opposite direction. 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 per cent 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 constant change in the dependent variable in the same direction, the 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 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 variables or from 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 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 variable y on the variable 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

$$\sum(y_i – a - bx_i)^2 \text{ with respect to a and b.}$$

The values of a and b are

$$a=y–bx$$

$$b= \frac{\sum xy - \frac{(\sum xi)(\sum yi)}{n}}{\sum xi^2 - \frac{(\sum x_i)^2}{n}}$$

The regression equation can now be written as

$$y=y–bx +bx$$
$$y - y =b(x - x)$$

Where b is the regression coefficient

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).

According to Drapper and Smith (1981), the method of stepwise multiple regression analysis is to insert variables in turn until the regression equation in 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 program, according to Burroughs Corporation (1975), first forms a correlation matrix, finds the best predictor (the independent variable having the highest correlation with criterion variable) and performs a regression analysis with this predictor. Then, the second best predictor (independent), 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 variable that is being considered for entrance into the regression equation is measured by the F-statistic. If F is too small (less than 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 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

ISBN: 978-81-930585-0-3

the specified tolerance level, which was 0.001 in the present example.

Path Analysis

The term 'path analysis' was first introduced by the biologist, Sewall Wright in 1934 in connection with decomposing the total correlation between any two variables in a causal 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. Path analysis makes use of standardized partial regression coefficient (known as beta weights) as effect coefficients. In linear additive effects are assumed, and then through path analysis a simple set of equations can be built 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 of paths: separate paths of influence leading through chronologically intermediate variable to which both the correlated variables have linked.

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

Factor Analysis

Factor analysis is a very useful and popular method of multivariate research technique, mostly used in social and behavioural sciences. This technique is applicable when there is a systematic interdependence among a set of observed or manifest variables, and the researcher is interested in finding out something more fundamental or latent which creates this communality (commonness). For example, we may have data on farmers' education, occupation, land, house, farm power, material possession, social participation etc. and want to infer from these some factor relating to social status, which shall summarize the communality of all the variables.

According to Kothari (1996), Factor analysis seeks to resolve a large set of measured variables in terms of relatively few categories, known as factors. This technique allows the researcher to group variables into factors (based on correlation between variables), and the factors so derived may be treated as new variables (often termed as latent variables) and their 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 variables is measured to obtain what are called factor loadings. Such factors loadings represent the correlation between the particular variable and the factor, and are usually placed in a matrix of correlations between the variable and the factors.

Concepts Used In Factor Analysis

Some important concepts used in factor analysis are explained, following Kothari (1996).

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

Communality (h^2) Communality, represented by h^2 , shows how much of each variable is accounted for by the underlying factor taken together. A high value of communality means that not much of the variable is left over after whatever the factors represent is taken into consideration.

Eigenvalue (or latent root): The sum of squared values of factor loadings relating to a factor is referred to as eigenvlue or latent root. Eigenvalue indicates the relative importance of each factor in accounting for the particular set of variables being analyzed.

Rotation: Rotation reveals different structures in the data and provides meaning to the results of factor analysis. There are different types of rotations such as orthogonal rotations, oblique rotations, varimax rotation etc. One has to select a rotation appropriate to the study. For the present study varimax rotation has been used.

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 ($\lambda_1 + \lambda_2 +$

$\cdots \ldots \ldots \ldots + \lambda_m$,) 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 ($\lambda^{1/2}.V$). 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 one.

Next step is to remove the noise imposed by (m - p) unnecessary axes. To accomplish this, p-orthogonal reference axes or factors are routed about the origin to positions such that the variance of the loading from each variable onto each factor axis is either extreme (±1) or zero. This maximization 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 insignificantly loadings. The column showing communality ($\sum h^2_j$) is the total amount of variance of each variable retained in the factors, and is computed by summing the squares of 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.

Canonical Correlation Analysis

A canonical correlation is the correlation of two canonical (latent) variables, one representing a set of independent variables the other a set of dependent variables. Each set may be considered a latent variable based on measured indicator variables in its set. The canonical correlation is optimized such that the linear correlation between the two latent variables is maximized. Whereas multiple regressions are used for many to one relationship, canonical correlation is used for many to many relationships. There may be more than one such linear correlation relating the two set of variables, with each such correlation representing a different dimension by which the independent set of variables is related to the dependent set. The purpose of canonical correlation is to explain the relation of the two sets of variables not to model the individual variables.

Analogous with ordinary correlation, canonical correlation squared is the per cent of variance in the dependent set of variables along a given dimension (there may be more than one). In addition to asking how strong the relationship is between two latent variables, canonical correlation is useful in determining how many dimensions are needed to account for that relationship. Canonical correlation finds the linear combination of variables that produces the largest correlation with second set of variables. This linear combination or "root" is extracted and the process is repeated for the residual data, with the constraint that the second linear combination of variables must not correlate with the first one. The process is repeated until a successive linear combination is no longer significant.

Canonical correlation is a member of multiple general linear hypothesis (MLGH) family and shares many of assumptions of multiple regression such as linearity of relationship, homoscedasticity (same level of relationship for the full range of the data), interval or near interval data, untruncated variables, proper specification of model, lack of high multicolinearity and multivariate normality for purpose of hypothesis testing.

Often in applied research, scientist encounter variables of large dimensions and are faced with problem of understanding dependency structures reduction of dimensionalities, construction of a subset of good predictors from the explanatory variables etc. Canonical correlation analysis provides us with a tool to attack these problems.

Some comments on the Canonical Correlation

There could be a situation where some of variables have high structure correlation even though their canonical weights are near zero. This could happen because the weight is partial coefficient whereas the structure correlations (canonical factor loading) are not: if a given variable share variance with other independent variable entered in the linear combinations of variables entered in the linear combinations of variables used to create a canonical variable, its canonical coefficient (weight) is computed based on the residual variance it can explain after controlling for these variables. If an independent variable is totally redundant with another independent variable, its partial coefficient (canonical weight) will be zero. Nonetheless, such a variable might have high correlation with canonical variable (that is high structures correlation) have to do with the sample overall correlation of the original variable with the canonical variable.

Canonical correlation is not a measure of the per cent of variance explained in the original variables. The square of the structure correlation is the per cent of variance in a given original variable accounted for by a given canonical variable on a given (usually the first) canonical correlation. Note that the average per cent of variance explained in the original variable by a canonical variable (the mean of the squared structure correlation for the canonical variable) is not all the same as the canonical correlation, which has to do with the

ISBN: 978-81-930585-0-3

correlation between the weighted sums of the two sets of variables. Put another way the canonical correlation does not tell us how much of the variance in the original variable is explained by the canonical variable instead, that is determined on the basis of the squares of the structure correlation.

Canonical coefficient can be used to explain with which original variable a canonical correlation is predominantly associated. The canonical coefficient are standardized coefficient and (like beta weights in regression). The magnitude can be compared. Looking at the columns in SPSS output which list the canonical coefficient as columns and the variable is a set of variables as rows, some researches simply not variable with the highest coefficient to determine which variable are associated with which canonical correlation and use this as the basis for inducing the meaning of the dimension represented by canonical correlations.

Redundancy In Canonical Correlation Analysis

Redundancy is the per cent of variance in one set of variable accounted for by the variate of the other set. The researcher wants high redundancy indicating that independent variate accounts for a high per cent of variance in the dependent set of original variables. Note that this is not the canonical correlation squared which the per cent of variance in the dependent variate is accounted for the independent variate.

Discriminant Function Analysis

Discriminant function analysis undertakes the same task as multiple linear regressions by predicting an outcome. However, multiple linear regressions is limited to cases where the dependent variable on the Y axis is an interval variable so that the combination of predictors will, through the regression equation, produce estimated mean population numerical Y values for given values of weighted combination of X values. But many interesting variables are categorical such as political party voting intensions, migrant/non-migrant status, making a profit or not, holding a particular credit card, owning, renting or paying a mortgage for a house, employed/unemployed, satisfied versus dissatisfied employees, which customers are likely to buy a product or not buy, what distinguishes stiller Bean clients from Gloria Beans clients, whether a person is a credit risk or not, etc.

Discriminant analysis is used when:

a) The dependent is categorized with the predictor IV's at interval level such as age, income, attitudes, perceptions and years of education, although dummy variables can be used as predictors as in multiple regressions logistic regressions IV's can be any level of measurement.
b) There are more than two DV categories, unlike logistic regression, which is limited to dichotomous dependent variables.

Discriminant Analysis Linear Equation

Discriminant analysis linear equation involves the determination of a linear equation like regression that will predict which group the case belongs to. The form of the equation or function is

$$D = V1x1 + V2x2 + V3x3 = \cdots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots Vixi + a$$

Where, D=Discrimination function;

V=The discriminant coefficient or weight for that variable.
X=Respondent's score for that variable.
a=a constant
i=the number of predictor variable
This function is similar to a regressions equation or function. The V's are unstandardized discriminant coefficient analogue to the 'b's' in the regressions equation. These V's maximize the distance between the mean of the criterion (dependent) variable. Standardized discriminant coefficient can also be used like beta weight in regressions. Good predictors tend to have large weights. What is needed, this function to do is maximize the distance between the categories; i.e. come up with an equation that has strong discriminatory power between groups. After using an existing set of data to calculate the discriminant function and classify cases, any new cases can then be classified. The number of discriminant function is one less the number of groups. There is only one function for the basic two group discriminant analysis.

A discriminant score is a weighted linear combination (sum) of the discriminating variables.

Assumptions of discriminant analysis:

a) The observations are random sample;
b) Each predictor variable is normally distributed;
c) Each of the allocations for the dependent categories in the initial classification are usually classified;
d) There must be at least two groups or categories in the initial classification are correctly classified;
e) There must be at least two groups or categories, with each case belonging to only one group so that the groups are mutually exclusive and collectively exhaustive (all cases can be placed in a group);
f) Each group or category must be well defined, clearly differentiated from any other (group) and natural. Putting a median split on an attitude scale is not a natural way to form group. Partitioning quantitative variables is only justified if there are easily identifiable gaps at the points of division.
g) For instance, three groups taking their available levels of amount of housing loan;
h) The groups or categories should be defined before collecting the data;

ISBN: 978-81-930585-0-3

i) The attributes used to separate the group should discriminate quite clearly between the groups so that group or category overlap is clearly non-existent or minimal;
j) Group sizes of the dependent should not be grossly different and should be at least five times the number of independent variables.

There are several purposes of Discriminant analysis;

i. To investigate differences between groups on the basis of the attributes of the cases, indicating which attributes contribute most to group separation. The descriptive technique successively identifies the linear combinations of attributes known as canonical discriminant functions (equation) which contribute maximally to group separation.
ii. Predictive discriminant analysis addresses the question of how to assign new cases to groups. The discriminant analysis function uses a person's scores on the predictor variables to predict the category to which the individual belongs.
iii. To determine the most parsimonious way to distinguish between groups.
iv. To classify cases into groups, statistical significance tests using chi-square enable you to see how well the function separates the groups.
v. To test theory whether cases are classified as predicted.
vi. Discriminate analysis creates an equation which will minimize the possibility of misclassifying cases into their respective groups or categories.

The aim of the statistical analysis in discriminant analysis is to combine (wieght) the variable scores in same way so that a single new composite variables, the discriminant score is produced. One way of thinking about this is in terms of a food recipe, where changing the proportions (weights) of the ingredients will change the characteristics of the finished cakes. Hopefully the weighted combinations of ingredients will produce two different types of cake.

Similarly, at the end of the discriminant process, it is hoped that each group will have a normal distribution of discriminant scores. The degree of overlap between the discriminant score distribution can then be used as a measure of the success of the technique, so that, like the different types of cake mix, we have two different types of groups

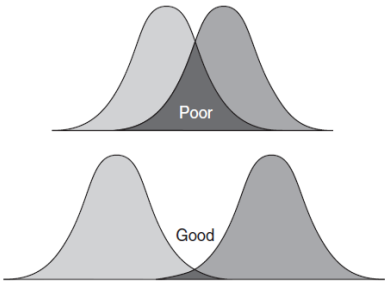

Fig. 4.2: Discriminant distribution

The top two distributions in figure overlap too much and do not discriminate too well compared to the bottom set. Misclassification will be minimized in the lower pair, whereas many will be misclassified in the top pair.

Standardizing the variables ensure that scale differences between the variables are eliminated. When all variables are standardized, absolute weights (i.e. ignore the sign) can be used to rank variables in terms of their discriminating power, the largest weight being associated with the most powerful discriminating variables with large weight are those which contribute mostly to differentially the groups.

As with most other multivariate methods, it is possible to present a pictorial explanation of the technique the following example uses a very simple data set, two groups and two variables. If scatter graphs are plotted for scores against the two variables, distributions like those in figure

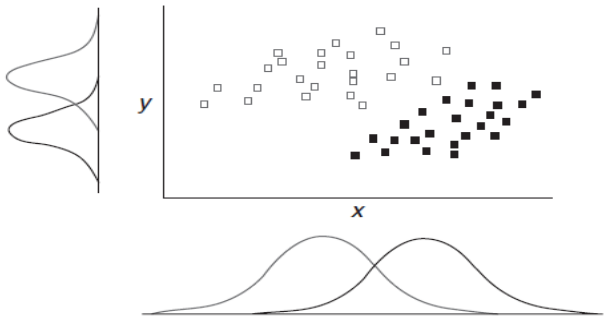

Fig. 4.3: Scattergraph displaying distribution

The new axis represents a new variable which is a linear combination of x and y i.e. it is a discriminant function (Fig. 4.3) obviously, with more than two groups or variables this graphical method becomes impossible.

ISBN: 978-81-930585-0-3

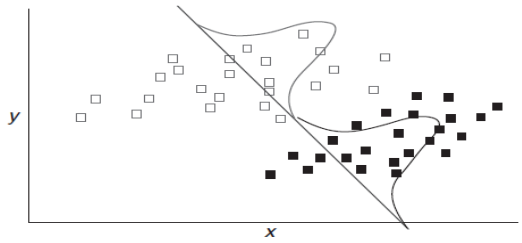

Fig. 4.4: New axis creating greater discrimination

Clearly, the two groups can be separated by their two variables, but there is a large amount of overlap on each single axis (although the y variable is the 'better' discriminator). It is possible to construct a new axis which passes through the two group centroids (means') such that the groups do not overlap on the new axis.

In a two-group situation predicted membership is calculated by first producing a score of D for each case using the discriminant function. Then cases with D values larger are classified into the other group. SPSS will save the predicted group membership and D scores as new variables.

The group centroid is the mean value of the discrimination score for a given category of the dependent variable. There are as many centroids as there are groups or categories. The cut-off is the mean of the centroids. If the discriminant score of the function is less than or equal to the cut-off the case is classed as 0, whereas if it is above, it is classed as 1.

Stepwise Discriminant Analysis

Discriminant analysis uses a collection of interval variables to predict a categorical variable that may be dichotomy or have more than two values. The technique involves finding a linear combination of independent variables (predictors)–the discriminant function- that creates the maximum difference between group memberships in the categorical dependent variable

In the present study for the stepwise discriminant analysis, canonical disrciminant function coefficients have been used. Stepwise discriminant analysis, like its parallel in multiple regressions, is an attempt to find the best set of predictors. It is often used in exploratory situation to identify those variables from among a large number that might be used later in a more rigorous theoretically driven study. In a stepwise discriminant analysis, the most correlated independent is entered first by the stepwise program, and then second until an additional dependent adds no significant amount to canonical R squared. The criteria of adding or removing are typically the setting of critical significance level for 'F' to remove. These are unstandardized coefficient (b) and used to create the discriminant function (equation). It operates just like the regression equation.

The discriminant function coefficient 'b' or standardized form 'beta' both indicates the partial contribution of each variable to the discriminant function controlling for all other variables in the equation. They can be used to asses each XI's (in the present study) unique contribution to the discriminant function and therefore provide information on the relative importance of each variable. If there are any dummy variables as in regression, individual 'beta' weights cannot be used and dummy variables must be assessed as a group through hierarchical discriminant analysis running the analysis, first without the dummy variables then with them. The difference is squared canonical correlation indicates the explanatory effect on the set of dummy variable.

Canonical Discriminant Function

Canonical Dicriminant Coefficient Table

The understanderdized coefficients (b) are used to create the dicriminant function (equation). It operates just like regression equation. The discriminant function coefficient b or standardized farm beta both indicates the partial contribution of each variable to the discriminant functions controlling for the discriminant function and therefore provide information on the relative importance of each variable. If there are any dummy variables as in regression, individual beta weights cannot be used and dummy variables must be assessed as a group through hierarchical discriminant analysis running the analysis first without the dummy variables then with them. The difference is squared canonical correlation indicates the explanatory effect on the set of dummy variable.

Group Centroid Table

A further way of interpreting discriminant analysis results is to describe each group in terms of profile using the group means of the predictor variables. These group means are called centroid. These are displayed in group centroid tables.
Wilks' Lamda table:
This table reveals that all the predictors add some predictive power to the discriminant function as all are significant with $p < .000$.